第一章　绪　　论

§1-1　估计与估计对象

什么是估计？简要地说，就是利用测量数据对未知变量或未知时间过程进行推算，推算的方法称为估计方法，推算的结果称为估计。若被估计对象是变量就称为变量估计，若被估计对象是时间过程就称为过程估计。若对时间过程在频率域上的谱分布进行估计则称为谱估计。

当用任何一种方法测量某些物理量（如长度、速度、重量、电压等）时，不可避免地具有随机误差，如进行多次测量，各次读数可能不相同，我们的任务就是要从这些测量值中对物理量作出尽可能好的估计，或者说希望估计值尽可能接近真值，这些物理量的真值是期望得到的，称之谓期望信号。

为了得到好的估计，人们发展了许多种估计方法，然而不论采用何种方法，也不可能得到真值，仍然具有一定的误差。由于采用的估计方法不同，估计的效果也不尽相同。因此需要建立统一标准对估计效果加以衡量，以便评出孰优孰劣。

此外，估计需要按照一定的准则进行，由于所采用准则的不同，将得出不同的估计方法。一种估计方法也只是在某种条件下适用，所以我们将探讨各种估计方法与所应用的准则及适用条件之间的关系。

下面对以上两个问题进行详细说明。

§1-2　估计质量评价

标量情况

我们用符号$\hat{x}$表示对未知物理量的一个估计，$\hat{x}$本身是一个随机变量，它具有一定形状的概率密度函数$p(\hat{x})$，我们希望$\hat{x}$集中在真值x的附近。例如图1-2-1中估计量1要优于估计量2，通常用下面三个指标来评价估计的质量。

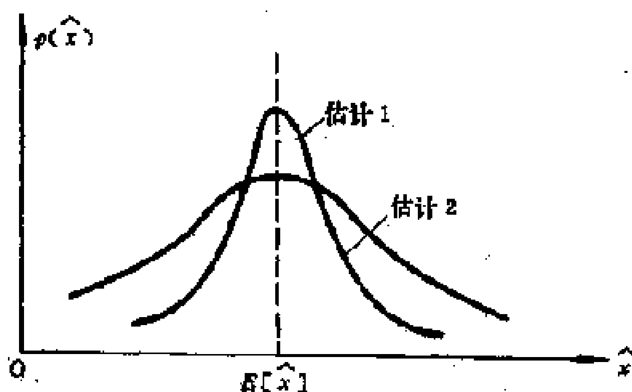


图1-2-1　两个标量估计的比较

1.　**偏量**

估计量的均值$E[\hat{x}]$与期望值$\overline{x}$之差定义为此估计值的偏量b，即有

$$b = E[\hat{x}] - \overline{x} = E[\hat{x} - \overline{x}] \tag{1-2-1}$$

当偏量b等于零时，称此估计$\hat{x}$为无偏估计。

若测量数据有限时的估计是有偏的，而当测量数据无限时，偏量趋于零，则称此估计为渐近无偏估计。

2. 方差

估计$\hat{x}$的方差定义为

$$\mathrm{var}[\hat{x}] = E[(\hat{x} - m)^2] = \sigma_{\hat{x}}^2 \tag{1-2-2}$$

$$m = E[\hat{x}]$$

假定$\hat{x}$的概率密度为高斯分布，即有

$$p(\hat{x}) = \frac{1}{\sigma_{\hat{x}}\sqrt{2\pi}} \exp\left[-\frac{(\hat{x} - m)^2}{2\sigma_{\hat{x}}^2}\right] \tag{1-2-3}$$

方差$\sigma_{\hat{x}}^2$决定了密度函数$p(\hat{x})$的分布宽度，它描述了估计值偏离其均值的散度，$\sigma_{\hat{x}}^2$越小，表明估计效果越好。

3. 均方误差

在许多情况下，具有较小偏量的估计却具有较大的方差，反之亦然。这一事实给对于两个估计的评价带来了困难，因此我们定义估计的均方误差为

$$MSE = E[(\hat{x} - \bar{x})^2] = b^2 + \sigma_{\hat{x}}^2 \tag{1-2-4}$$

在均方误差中同时包括偏量和方差。如果随着测量次数增多，使均方误差越来越小，即偏量与方差都趋于零，则称此估计为**一致性估计**

若估计是无偏的，则均方误差就等于方差，均方误差最小与方差最小等价。

矢量情况

例如要估计平面上一点的坐标位置，这是一个二维矢量的估计问题，此二维矢量为

$$\hat{X} = \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} \tag{1-2-5}$$

这时估计偏量是一矢量

$$\boldsymbol{b} = \begin{bmatrix} \hat{x}_1 - E[\hat{x}_1] \\ \hat{x}_2 - E[\hat{x}_2] \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{1-2-6}$$

这时需要用估计误差的协方差矩阵来表达估计的分散度，此协方差矩阵定义为

$$\boldsymbol{R} = E[(\hat{X} - \boldsymbol{m})(\hat{X} - \boldsymbol{m})^T] \tag{1-2-7}$$

其中 $\boldsymbol{m} = E[\hat{X}]$

为了表达估计的分散度，需要研究误差$\hat{x}_1$、$\hat{x}_2$的联合分布密度。为了简化代数运算，不失一般性，令$\boldsymbol{m} = 0$，则$\hat{x}_{1*}$、$\hat{x}_{2*}$的联合概率密度函数为

$$p(\hat{X}_*) = \frac{1}{2\pi|\boldsymbol{R}|^{1/2}} \exp\left[-\frac{1}{2}\hat{X}_*^T \boldsymbol{R}^{-1} \hat{X}_*\right] \tag{1-2-8}$$

其中 $\hat{X}_* = \begin{bmatrix} \hat{x}_{1*} \\ \hat{x}_{2*} \end{bmatrix}$

记
$$\hat{X}_r^T R^{-1} \hat{X}_r = c^2,\quad c\text{为常数} \tag{1-2-9}$$
展开(1-2-9) 式，得到一椭圆方程，它是二维概率密度函数的一个剖面，现在来求误差矢量$\hat{X}_r$落在此椭圆内的概率。椭圆的面积为

$$A = |R|^{\frac{1}{2}}\pi c^2 \tag{1-2-10}$$

与 c 和 $c+\delta c$ 相对应的两个椭圆之间的微分面积为

$$\delta A = |R|^{\frac{1}{2}} 2\pi c\,\delta c \tag{1-2-11}$$

这个微分面积上概率密度的高度为

$$(2\pi|R|^{\frac{1}{2}})^{-1}\exp\left(-\frac{c^2}{2}\right) \tag{1-2-12}$$

用(1-2-11)乘(1-2-12)式，并进行积分。记落入椭圆内的概率为P，则在椭圆外的概率为$1-P$。

$$1-P=\int_c^\infty u\exp\left(-\frac{u^2}{2}\right)du=\exp\left(-\frac{c^2}{2}\right) \tag{1-2-13}$$

$$P=1-\exp\left(-\frac{c^2}{2}\right) \tag{1-2-14}$$

椭圆方程$\hat{X}_r^T R^{-1}\hat{X}_r=c^2$称为不确定椭圆或密集椭圆，如图1-2-2所示。

对于任意k维矢量，类似结果也成立，我们将在$k+1$维空间中得到一个多维空间的超椭球。

对于图1-2-2中的不确定椭圆，当$\hat{x}_{1r}$和$\hat{x}_{2r}$不相关时，R^{-1}是对角线阵，不确定椭圆是一正椭圆，而当$\hat{x}_{1r}$和$\hat{x}_{2r}$相关时，R^{-1}是一方阵，不确定椭圆是一斜椭圆。

图1-2-2 二维高斯密度函数图形

在矢量情况，均方误差变成了均方程差矩阵P。

$$P=E[(\hat{X}-X)\times(\hat{X}-X)^T]$$

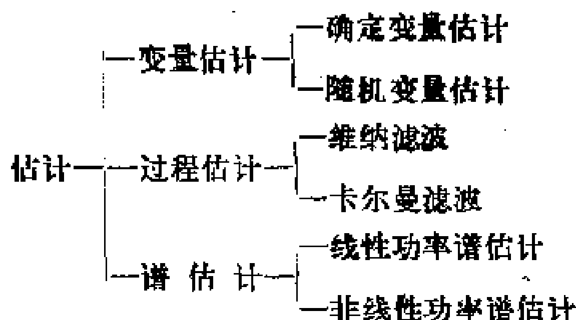
通常用P的迹来表达矢量估计的均方误差。

$$\begin{aligned}\mathrm{tr}P&=\mathrm{tr}E[(\hat{X}-X)(\hat{X}-X)^T]\\&=E[(\hat{X}-X)^T(\hat{X}-X)]\end{aligned}$$

当测量数据无限增多时，$\mathrm{tr}P$趋于零，则称此矢量估计为一致性估计。倘若矢量估计是无偏的，则均方误差矩阵就等于协方差矩阵。

§1-3 估计方法名称、符号，应用准则与条件

本书中把各种估计按下表加以分类：



所谓变量估计是这样一类估计问题，根据测量值集合对期望值（标量或矢量）进行估算。当期望值是确定变量时，就称为确定变量估计，当期望值是随机变量时，就称为随机变量估计。

所谓过程估计是根据测量时间序列对期望时间过程进行估算。过程估计方法分成维纳滤波与卡尔曼滤波两大类。维纳滤波是建立在批量数据处理基础之上的，以横向滤波器作为其基本结构形式；卡尔曼滤波是建立在递推数据处理基础之上的，以递归滤波器作为其基本结构形式。

所谓功率谱估计是对随机时间信号在频率域上的功率谱分布进行推算。功率谱估计方法也分成两大类。一类是线性功率谱估计，它是建立在线性付立叶变换的基础之上的；另一类是非线性功率谱估计，它是建立在信号模型辨识的基础之上的。由于谱分布与时间序列之间成非线性变换关系，故称为非线性功率谱估计。

下面对每一种估计方法的名称、符号，应用的准则与使用条件以表格形式列出，本节所述实际上是全书的纲要，也是对各种估计方法的一个归纳。

表1-1 确定变量估计

估计名称	符号	准则	先验统计知识
最小二乘估计	$\hat{X}_{ls}$	平方误差最小	不需要
最优加权最小二乘估计	$\hat{X}_{lsw}$	加权平方误差最小	测量噪声协方差阵
最大似然估计	$\hat{X}_{ml}$	条件密度 $P(Z/X)$ 最大	条件密度 $P(Z/X)$

其中 $Z=[z_1,z_2,\cdots,z_N]^T$ 表示测量集合

表1-2 随机变量估计

估计名称		符号	准则	先验统计知识
贝叶斯估计	最大后验估计	$\hat{X}_{map}$	后验密度 $P(X/Z)$ 最大	$P(X/Z)$ 已知
	最小方差估计	$\hat{X}_{mv}$	后验密度 $P(X/Z)$ 的均值	$P(X/Z)$ 已知
线性最小方差估计		$\hat{X}_{lmv}$	均方误差最小 $\hat{X}_{lmv}=A+BZ$	Z和X的头两阶矩
最小均方误差估计		$\hat{X}_{lms}$	均方误差最小 $\hat{X}_{lms}=W^TZ$	自相关矩阵 ϕ_{zz} 互相关矩阵 ϕ_{zx}

表1-3 随机时间过程估计

估计名称	准则	先验统计知识
维纳滤波 （自适应维纳滤波）	均方误差最小	自相关矩阵ϕ_{xx} 互相关矢量ϕ_{zx}
卡尔曼滤波 （自适应卡尔曼滤波）	线性无偏最小方差	状态噪声协方差阵Q_k 测量噪声协方差阵R_k

表1-4 功率谱估计

名称	方法
线性功率谱估计	相关函数法 周期图法
非线性功率谱估计	时间序列的参数估计法

第一篇 变量估计

第二章 确定变量估计

我们从最基本的估计问题开始，限定被估计量是确定变量。如果对一确定变量进行多次测量，由于测量噪声的存在，各次测量值不相同。因此，需要建立起一套算法，从这些测量值得出确定变量的合理估计。

本章第一节讨论经典的最小二乘估计，第二节讨论最优加权最小二乘估计，后者是前者的改进。第三节讨论非线性最小二乘估计，是线性最小二乘估计在非线性问题中的推广。第四节讨论最大似然估计，最大似然估计通常看成是下章中贝叶斯估计的一个特例，其估计对象仍是确定变量，因而把它列人本章之中。

接着讨论最小二乘数据拟合，所谓数据拟合就是用一确定的数学表达式来近似表示某随机变量与另一确定变量之间的关系，它实际上是最小二乘估计方法的一种应用。最后讨论有约束条件时的最小二乘估计。

§2-1 最小二乘估计$\hat{X}_{ls}$

最小二乘估计方法是一种最早的估计方法，二百多年前由高斯提出。

标量情况

为了估计未知值x，对它进行了N次测量，测量值是

$$z_i = h_i x + e_i \qquad i = 1, 2, \cdots, N \tag{2-1-1}$$

其中h_i为已知常数，e_i为第i次测量误差。设所求的估计值为$\hat{x}$，则第i次观测值与相应的估计值$h_i\hat{x}$之间的误差为

$$\hat{e}_i = z_i - h_i\hat{x} \tag{2-1-2}$$

将N次误差的平方之和记作

$$J(\hat{x}) = \sum_{i=1}^{N}(z_i - h_i\hat{x})^2 \tag{2-1-3}$$

能使$J(\hat{x})$取极小值的估计值$\hat{x}$就称为未知量x的最小二乘估计，记作$\hat{x}_{ls}$。使$J(\hat{x})$取极小就称为最小二乘准则，根据最小二乘准则求估计值的方法就称为最小二乘估计方法。

采用矩阵符号，可使求最小二乘估计的表达式简练。记

$$Z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{pmatrix}, \qquad H = \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{pmatrix}, \qquad e = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{pmatrix} \tag{2-1-4}$$

把(2-1-1)、(2-1-3)式改写成

$$Z = Hx + e \tag{2-1-5}$$

$$J(\hat{x}) = (Z - H\hat{x})^T(Z - H\hat{x}) \tag{2-1-6}$$

求$J(\hat{x})$的极小值，令

$$\frac{\partial J(\hat{x})}{\partial \hat{x}} = -2H^T(Z - H\hat{x}) = 0 \tag{2-1-7}$$

解出最小二乘估计$\hat{x}_{LS}$

$$\hat{x}_{LS} = (H^TH)^{-1}H^TZ \tag{2-1-8}$$

由于 $H^TH = h_1^2 + h_2^2 + \cdots + h_N^2 > 0$

显见

$$\frac{\partial^2}{\partial \hat{x}^2}J(\hat{x}) = 2H^TH > 0 \tag{2-1-9}$$

证明了$\hat{x}_{LS}$是使$J(\hat{x})$达到极小值。

当$h_1 = h_2 = \cdots = h_N = 1$时，容易算出

$$\hat{x}_{LS} = \frac{1}{N}\sum_{i=1}^{N} z_i \tag{2-1-10}$$

这是N次测量值的算术平均值，所以，算术平均是最小二乘估计的一个特例。

$\hat{x}_{LS}$既然是一个估计量，它必然是一个随机量，存在着方差。因为估计误差为

$$\begin{aligned} x - \hat{x}_{LS} &= (H^TH)^{-1}(H^TH)x - (H^TH)^{-1}H^TZ \\ &= (H^TH)^{-1}H^T(Hx - Z) \\ &= -(H^TH)^{-1}H^Te \end{aligned} \tag{2-1-11}$$

假设测量误差矢量e的均值为零，方差阵为R，即

$$E[e] = 0 \qquad E[ee^T] = R$$

则有

$$E[x - \hat{x}_{LS}] = -(H^TH)^{-1}H^TE[e] = 0$$

可见，当测量误差为零均值时，最小二乘估计为无偏估计。

此时估计误差的方差为

$$\begin{aligned} P_{LS} &= E[(x - \hat{x}_{LS})(x - \hat{x}_{LS})^T] \\ &= E[(H^TH)^{-1}H^T(ee^T)H(H^TH)^{-1}] \\ &= (H^TH)^{-1}H^TRH(H^TH)^{-1} \end{aligned} \tag{2-1-12}$$

若各测量误差独立，且方差相等，即$\mathrm{var}\,e_i = \sigma^2$，则得出

$$P_{LS} = (H^TH)^{-1}\sigma^2 \tag{2-1-13}$$

当$h_1 = h_2 = \cdots = h^N = 1$时

$$P_{LS} = \frac{1}{N}\sigma^2 \tag{2-1-14}$$

显见，当测量次数N增多时，方差P_{LS}趋近于零。所以，当各测量误差为零均值时，最小二乘估计为无偏一致性估计。

例2-1-1　设用两个电压表测量某电压，第一个表的读数是220伏，已知该表的方差为

8^2（伏^2），第二个表的读数是210伏，已知该表的方差为4^2（伏^2），求估计值$\hat{x}_{LS}$及其方差。

解：

$$\hat{x}_{LS}=\frac{220+210}{2}=215\text{伏}$$

$$P_{LS}=\frac{1}{2}[1 \quad 1]\begin{bmatrix}8^2 & 0\\ 0 & 4^2\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix}\frac{1}{2}$$

$$=\frac{1}{4}(8^2+4^2)=20\text{伏}^2$$

矢量情况

假设被估计量X是n维矢量，测量Z为m维矢量，则测量矩阵H应是$m\times n$维矩阵，可得出测量方程为

$$\underset{m\times 1}{Z}=\underset{m\times n}{H}\ \underset{n\times 1}{X}+\underset{m\times 1}{\varepsilon} \tag{2-1-15}$$

仿照标量情况，容易求出最小二乘估计及其方差阵，(2-1-8)式和(2-1-12）式仍然适用，但改成矢量符号

$$\hat{X}_{LS}=(H^TH)^{-1}H^TZ \tag{2-1-16}$$

$$P_{LS}=(H^TH)^{-1}H^TRH(H^TH)^{-1} \tag{2-1-17}$$

递推最小二乘估计

如果我们已根据N次测量值求出了最小二乘估计$\hat{X}_{LS,N}$及其方差$P_{LS,N}$，现在又新增加了一个测量值，必须重新计算一遍才能得出新的估计及其方差，这就需要存储所有测量数据，而且计算量很大，十分不便。因此希望找到一种递推算法，可以由$\hat{X}_{LS,N}$和$P_{LS,N}$，加上新测量值直接推算出$\hat{X}_{LS,(N+1)}$和$P_{LS,(N+1)}$。现在来推导这种递推算法。

在有N个测量值时，记测量方程为

$$Z_N=H_NX+\varepsilon_N \tag{2-1-18}$$

X的最小二乘估计为

$$\hat{X}_{LS,N}=(H_N^TH_N)^{-1}H_N^TZ_N \tag{2-1-19}$$

现在又得到了一个新测量值

$$Z_{N+1}=h_{N+1}X+\varepsilon_{N+1} \tag{2-1-20}$$

把(2-1-20)式加到(2-1-18)式中去，得到新的测量方程为

$$Z_{N+1}=H_{N+1}X+\varepsilon_{N+1} \tag{2-1-21}$$

其中

$$Z_{N+1}=\begin{bmatrix}Z_N\\ \vdots\\ Z_{N+1}\end{bmatrix} \qquad H_{N+1}=\begin{bmatrix}H_N\\ h_{N+1}\end{bmatrix} \qquad \varepsilon_{N+1}=\begin{bmatrix}\varepsilon_N\\ \varepsilon_{N+1}\end{bmatrix}$$

由$N+1$个测量值得到的估计应是

$$\hat{X}_{LS,(N+1)}=(H_{N+1}^TH_{N+1})^{-1}H_{N+1}^TZ_{N+1} \tag{2-1-22}$$

为了方便起见，记

$$P_N=(H_N^TH_N)^{-1} \quad \text{或} \quad P_N^{-1}=(H_N^TH_N)$$

于是有

$$\hat{X}_{L,N}=P_N H_N^T Z_N \tag{2-1-23}$$

$$\hat{X}_{L,(N+1)}=P_{N+1}H_{N+1}^T Z_{N+1} \tag{2-1-24}$$

其中

$$P_{N+1}=(H_{N+1}^T H_{N+1})^{-1}=\left\{[H_N^T h_{N+1}^T]\begin{bmatrix}H_N\\ h_{N+1}\end{bmatrix}\right\}^{-1}$$

$$=[P_N^{-1}+h_{N+1}^T h_{N+1}]^{-1} \tag{2-1-25}$$

利用附录中矩阵等式

$$(P^{-1}+H^T R^{-1}H)^{-1}=P-PH^T(HPH^T+R)^{-1}HP$$

可以得出

$$P_{N+1}=P_N-P_N h_{N+1}^T(h_{N+1}P_N h_{N+1}^T+1)^{-1}h_{N+1}P_N \tag{2-1-26}$$

可见，由已知P_N就可算出P_{N+1}。再看

$$\hat{X}_{L,(N+1)}=P_{N+1}H_{N+1}^T Z_{N+1}=P_{N+1}\begin{bmatrix}H_N\\ h_{N+1}\end{bmatrix}^T\begin{bmatrix}Z_N\\ z_{N+1}\end{bmatrix}$$

$$=P_{N+1}[H_N^T Z_N+h_{N+1}^T z_{N+1}] \tag{2-1-27}$$

现在我们来研究(2-1-27)式右端第一项，利用(2-1-23)式，可得出

$$P_{N+1}H_N^T Z_N=P_{N+1}P_N^{-1}\hat{X}_{L,N} \tag{2-1-28}$$

再利用(2-1-25)式，得

$$P_N^{-1}=P_{N+1}^{-1}-h_{N+1}^T h_{N+1} \tag{2-1-29}$$

把(2-1-29)式代入(2-1-28)式中，得

$$P_{N+1}H_N^T Z_N=P_{N+1}[P_{N+1}^{-1}-h_{N+1}^T h_{N+1}]\hat{X}_{L,N}$$

$$=\hat{X}_{L,N}-P_{N+1}h_{N+1}^T h_{N+1}\hat{X}_{L,N}$$

记

$$K_{N+1}=P_{N+1}h_{N+1}^T \tag{2-1-30}$$

(2-1-27)式右边第二项可写成

$$P_{N+1}h_{N+1}^T z_{N+1}=K_{N+1}z_{N+1} \tag{2-1-31}$$

把(2-1-30)、(2-1-31)式代入(2-1-27)式，得

$$\hat{X}_{L,(N+1)}=(I-K_{N+1}h_{N+1})\hat{X}_{L,N}+K_{N+1}z_{N+1}$$

$$=\hat{X}_{L,N}+K_{N+1}[z_{N+1}-h_{N+1}\hat{X}_{L,N}] \tag{2-1-32}$$

用P_{N+1}左乘(2-1-29)式，得

$$P_{N+1}P_N^{-1}=I-K_{N+1}h_{N+1}$$

或

$$P_{N+1}=(I-K_{N+1}h_{N+1})P_N \tag{2-1-33}$$

对照(2-1-26)式，得出

$$K_{N+1}=P_N h_{N+1}^T[h_{N+1}P_N h_{N+1}^T+1]^{-1} \tag{2-1-34}$$

(2-1-32)、(2-1-33)、(2-1-34)式构成了$\hat{X}_{L,(N+1)}$的递推估计公式．但这里$P_N=(H^T H)^{-1}$，并不代表方差。

若(2-1-12)式中的$R=\sigma^2 I$，于是有

$$P_N=\sigma^2(H^T H)^{-1} \tag{2-1-35}$$

则有

$$K_{N+1}=P_N h_{N+1}^T[h_{N+1}P_N h_{N+1}^T+\sigma^2 I]^{-1} \tag{2-1-36}$$

相应的

$$P_{N+1}=[1-K_{N+1}h_{N+1}]P_N \tag{2-1-37}$$

$$\hat{X}_{LS(N+1)}=\hat{X}_{LSN}+K_{N+1}[Z_{N+1}-h_{N+1}\hat{X}_{LSN}] \tag{2-1-38}$$

(2-1-34)、(2-1-37)、(2-1-38)组成最小二乘估计递推公式。就上述递推公式需要说明三个问题：

1. 从(2-1-38)式可以看出，新估计$\hat{X}_{LS(N+1)}$是由老估计$\hat{X}_{LSN}$加上校正项而构成的，而校正项是与新测量值Z_{N+1}与$h_{N+1}\hat{X}_{LSN}$之差成比例的，其中$h_{N+1}\hat{X}_{LSN}$可看成对于Z_{N+1}的预估量，当预估正确时，校正项为零，当预估不准确时，就进行自动校正。我们把$(Z_{N+1}-h_{N+1}\hat{X}_{LSN})$称为新息，意思是说，测量值$Z_{N+1}$带来了新的信息，因而需要修改老的估计$\hat{X}_{LSN}$。

2. 在进行递推估计时，我们需要一组初始值$\hat{X}_0$和P_0，这可以根据先验知识选择一组初始值，如果没有任何先验知识可供参考的话，可以令$\hat{X}_0=0$，$P_0=c^2 I$以后将证明，只要c^2足够大，经过若干次递推之后，这种不准确的初始值的影响就会逐渐消失，最终能得到满意的估计值，之所以如此，是因为估计具有自动校正的性质。下面举例说明之。

例2-1-2 设对未知标量x进行了六次测量，测量方程为

$$\begin{pmatrix}3\\5\\4\\15\\11\\13\end{pmatrix}=\begin{pmatrix}2\\2\\2\\4\\4\\4\end{pmatrix}x+\begin{pmatrix}e_1\\e_2\\e_3\\e_4\\e_5\\e_6\end{pmatrix}$$

已知各e_i独立，$E[e_i]=0$，$\mathrm{var}[e_i]=1$

求最小二乘估计及其方差。

解：利用递推公式(2-1-34)、(2-1-37)、(2-1-38)递推计算P_i与$\hat{x}_i$。

设初始值 $\hat{x}_0=0$，$P_0=\infty$

今将计算结果列表如下，从计算结果可以看出，P_i是逐渐减小的，随着测量次数增多，P_i将趋近于零。

i	0	1	2	3	4	5	6
P_i	∞	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{1}{28}$	$\frac{1}{44}$	$\frac{1}{60}$
$\hat{x}_i$	0	$\frac{3}{2}$	2	2	3	$\frac{32}{11}$	3

3. 利用公式$\hat{X}_{LS}=(H^TH)^{-1}H^TZ$求最小二乘估计时，不需要任何先验统计知识，应用起来十分方便。若已知各测量噪声是零均值的，则所得估计$\hat{X}_{LS}$是无偏估计。若已知测量噪声的协方差阵R，则可以应用公式

$$P_{LS}=(H^TH)^{-1}H^TRH(H^TH)^{-1}$$

求得估计误差的协方差矩阵。

反之，若我们对测量噪声缺乏先验统计知识，那么就不能对估计$\hat{X}_{LS}$的质量作出评价。

§2-2 最优加权最小二乘估计$\hat{X}_{lsw}$

在第一章中，我们采用的准则是使误差平方和

$$J(\hat{X})=\sum_{i=1}^{n}(z_i-h_i\hat{X})^2 \tag{2-2-1}$$

达到最小。这意味着对每次测量都是同等看待的。这自然产生一个问题，由于每次测量误差不一定有相同的方差，采用如此准则是否合理？以例2-1-1来说，第一个表的方差是8^2，第二个表的方差是4^2，当然第二次测量比较准确，因此理应把第二次测量看得重些。极端地说，如果第二次测量误差的方差是零的话，那么只要利用第二次测量值就可以了，用不着再进行估计。由此想到，应当对各误差加上适当的权。采用的准则是使

$$J_w(\hat{X})=\sum_{i=1}^{N}W_i(Z_i-h_i\hat{X})^2 \tag{2-2-2}$$

达到最小。写成矩阵形式是

$$J_w(\hat{X})=(Z-H\hat{X})^TW(Z-H\hat{X}) \tag{2-2-3}$$

其中W为适当选取的对角线权矩阵。

现在来求使$J_w(\hat{X})$达到最小的估计$\hat{X}_{lsw}$，$\hat{X}_{lsw}$称为加权最小二乘估计。

利用矩阵微分法，对(2-2-3)式求导数，可得

$$\frac{\partial}{\partial\hat{X}}J_w(\hat{X})=-H^T(W+W^T)(Z-H\hat{X})$$

$$=-2H^TW(Z-H\hat{X}) \tag{2-2-4}$$

令上式等于零，解出

$$\hat{X}_{lsw}=(H^TWH)^{-1}H^TWZ \tag{2-2-5}$$

估计误差为

$$\begin{aligned}\tilde{X}&=X-\hat{X}_{lsw}\\&=(H^TWH)^{-1}H^TW(HX-Z)\\&=-(H^TWH)^{-1}H^TW\varepsilon\end{aligned} \tag{2-2-6}$$

若测量误差矢量均值为零，方差阵为R，即

$$E[\varepsilon]=0 \qquad E[\varepsilon\varepsilon^T]=R$$

则有

$$E[X-\hat{X}_{lsw}]=0$$

故$\hat{X}_{lsw}$为无偏估计。

估计误差的方差阵为

$$\begin{aligned}P&=E[\tilde{X}\tilde{X}^T]\\&=(H^TWH)^{-1}H^TWRWH(H^TWH)^{-1}\end{aligned} \tag{2-2-7}$$

现在的问题是要选择适当的加权阵W使P达到最小。若已知

$$R=\begin{pmatrix}\sigma_1^2 & & 0\\ & \sigma_2^2 & \\ & & \ddots \\ 0 & & \sigma_n^2\end{pmatrix} \qquad W=\begin{pmatrix}w_1 & & 0\\ & w_2 & \\ & & \ddots \\ 0 & & w_n\end{pmatrix}$$

不失一般性，令$H^T=[1, 1, \cdots, 1]$，则有

$$H^TWH=\sum_{i=1}^{n} w_i$$

$$H^TWRWH=\sum_{i=1}^{n} w_i^2\sigma_i^2$$

得

$$P=\frac{\sum_{i=1}^{n} w_i^2\sigma_i^2}{\left(\sum_{i=1}^{n} w_i\right)^2} \tag{2-2-8}$$

对W求导数并使之等于零。

$$\frac{dP}{dW}=\frac{1}{\left(\sum_{i=1}^{n} w_i\right)^2}\begin{pmatrix}2\sigma_1^2 w_1\\2\sigma_2^2 w_2\\\vdots\\2\sigma_n^2 w_n\end{pmatrix}$$

$$-\left(\sum_{i=1}^{n} w_i^2\sigma_i^2\right)\frac{2}{\left(\sum_{i=1}^{n} w_i\right)^3}\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}=0 \tag{2-2-9}$$

因为$\left(\sum_{i=1}^{n} w_i\right)^2\neq 0$，故有

$$\sigma_i^2 w_i=\frac{\sum_{i=1}^{n} w_i^2\sigma_i^2}{\sum_{i=1}^{n} w_i} \qquad i=1, 2, \cdots, n.$$

显然只有选择$w_i=\sigma_i^{-2}$才能满足上述等式，也就是当选择$W=R^{-1}$时，P达到最小，记作P_{lsw}，此时的估计记作$\hat{X}_{lsw}$，称为最优加权最小二乘估计。

$$\hat{X}_{lsw}=(H^TR^{-1}H)^{-1}H^TR^{-1}Z \tag{2-2-10}$$

$$P_{lsw}=(H^TR^{-1}H)^{-1} \tag{2-2-11}$$

由以上讨论可知，$\hat{X}_{lsw}$是一组独立无偏测量的线性组合，并具有无偏最小方差的性质，故$\hat{X}_{lsw}$是一种线性无偏最小方差估计。

推广到相关噪声情况

若各次测量噪声不独立，则噪声矢量的协方差阵为非对角线阵，这时可证明(2-2-5)与(2-2-7)式仍然成立。重写如下：

$$\hat{X}_{lsw}=(H^TWH)^{-1}H^TWZ$$

$$P_{lsw}=(H^TWH)^{-1}H^TWRWH(H^TWH)^{-1}$$

其中W为一个适当选取的对称正定加权矩阵。现在的问题是选取什么样的加权W阵才能使P_{lsw}达到最小。

这里先证明一个"矩阵型"许瓦兹不等式。设A、B分别为$n\times m$和$m\times l$矩阵，且AA^T为可逆阵，则有

$$
\begin{aligned}
0 &\leqslant [B-A^T(AA^T)^{-1}AB]^T[B-A^T(AA^T)^{-1}AB] \\
&= B^TB-2B^TA^T(AA^T)^{-1}AB \\
&\quad +B^TA^T(AA^T)^{-1}AA^T(AA^T)^{-1}AB \\
&= B^TB-(AB)^T(AA^T)^{-1}(AB)
\end{aligned}
\tag{2-2-12}
$$

得施瓦兹不等式如下

$$B^TB\geqslant(AB)^T(AA^T)^{-1}(AB) \tag{2-2-13}$$

因为R为对称正定矩阵，可表示为

$$R=C^TC \tag{2-2-14}$$

这里C为可逆阵，代入(2-2-7)式，有

$$P_{lsw}=[CWH(H^TWH)^{-1}]^T[CWH(H^TWH)^{-1}] \tag{2-2-15}$$

令　$A=H^TC^{-1}$，$B=CWH(H^TWH)^{-1}$

则有

$$
\begin{aligned}
P_{lsw} &= B^TB\geqslant(AB)^T(AA^T)^{-1}(AB) \\
&=[H^TC^{-1}CWH(H^TWH)^{-1}]^T[H^TC^{-1}(C^T)^{-1}H]^{-1} \\
&\quad \cdot[H^TC^{-1}CWH(H^TWH)^{-1}] \\
&=(H^TR^{-1}H)^{-1}
\end{aligned}
\tag{2-2-16}
$$

显然，上述不等式取等号时P_{lsw}达到最小，所以最佳加权矩阵应取

$$W=R^{-1} \tag{2-2-17}$$

例2-2-1　求例2-1-1的最优加权最小二乘估计。

解：取$W=R^{-1}=\begin{bmatrix}8^{-2} & 0\\ 0 & 4^{-2}\end{bmatrix}$

由(2-2-10)、(2-2-11)式求出

$$P_{lsw}=\frac{1}{8^{-2}+4^{-2}}=\frac{64}{5}=12.8\text{伏}^2$$

$$\hat{X}_{lsw}=\frac{64}{5}(220\times8^{-2}+210\times4^{-2})=212\text{伏}$$

比较例2-1-1和例2-2-1结果，$P_{ls}=20$伏^2，而$P_{lsw}=12.8$伏^2。

可见，最优加权最小二乘估计的质量要优于一般最小二乘估计。但是，求最优加权最小二乘估计时需要有关于测量误差方差阵R的先验知识，当没有关于测量噪声的先验知识时，就只能采用一般最小二乘估计方法。

研究(2-2-10)、(2-2-11)式。若各测量误差v_i互相独立，零均值且方差相等，即在平稳白噪声条件下，有

$$R=\sigma^2 I$$

代入(2-2-10)、(2-2-11)式中，得

$$\hat{X}_{lsw}=(H^TH)^{-1}H^TZ=\hat{X}_{ls} \tag{2-2-18}$$

$$P_{lsw}=(H^TH)^{-1}\sigma^2=P_{ls} \tag{2-2-19}$$

这时，两种估计结果相同。

所以，当已知关于各测量噪声方差时，应当采用最优加权最小二乘估计，估计的质量要好些，当没有这些知识时，就只好采用一般最小二乘估计，虽然估计质量差些，但也只好如此。

递推最优加权最小二乘估计

采用上节中推导递推最小二乘估计相类似的方法，可以得出最优加权最小二乘估计的递推公式，这里不加推导，列出结果如下：

若根据前 k 次测量已求得估计 $\hat{X}_k$ 和方差 P_k，现在又得到第 $k+1$ 次测量

$$z_{k+1}=h_{k+1}X_{k+1}+V_{k+1}$$

则可以递推地计算 $\hat{X}_{k+1}$ 和 P_{k+1}。递推公式为

$$\hat{X}_{k+1}=\hat{X}_k+K_{k+1}[z_{k+1}-h_{k+1}\hat{X}_k] \tag{2-2-20}$$

$$P_{k+1}=[1-K_{k+1}h_{k+1}]P_k \tag{2-2-21}$$

$$K_{k+1}=P_kh_{k+1}^T[h_{k+1}P_kh_{k+1}^T+R_{k+1}]^{-1} \tag{2-2-22}$$

其中 $R_{k+1}=E[V_{k+1}V_{k+1}^T]$，$E[V_{k+1}]=0$

给定初始值 $\hat{X}_0$ 和 P_0，就可根据每一个测量值递推地求出 $\hat{X}_k$ 和 P_k。

上述递推最优加权最小二乘估计公式(2-2-20)、(2-2-21)、(2-2-22)式可以应用于参数 h_k、R_k 随时间变化的情况。递推与非递推最优加权最小二乘估计是等价的，但递推公式应用起来要方便些。

阻尼最小二乘估计 $\hat{X}_{LSd}$

对于噪声方差 R_k 随时间变化的情况，用普通递推最小二乘法求估计显然是不相宜的，因为在那里已假定噪声方差是时不变的，而用递推最优加权最小二乘估计时又需要关于 R_k 的先验知识，实际上往往不具备，为此提出了一种次优最小二乘估计方法。

为了体现参数的时变性质，应强调最近观测数据的作用，而渐渐遗忘陈旧数据的影响。

重写(2-1-6)式，普通最小二乘估计的目标函数为

$$J(\hat{X})=(Z-H\hat{X})^T(Z-H\hat{X})$$

$$=\sum_{i=1}^{k}(z_i-h_i\hat{X})^2 \tag{2-2-23}$$

现在令

$$J(\hat{X})=\sum_{i=1}^{k}\lambda^{k-i}(z_i-h_i\hat{X})^2 \qquad 0<\lambda<1 \tag{2-2-24}$$

也就是对求和的每一项乘了一个加权系数 λ^{k-i}，λ 叫做遗忘因子，这体现了在目标函数中强调了新近数据的作用。

对目标函数取极小化，令 $\frac{\partial J}{\partial \hat{X}}=0$。类似地可以求得递推阻尼最小二乘估计公式

$$\hat{X}_{k+1}=\hat{X}_k+K_{k+1}[z_{k+1}-h_{k+1}\hat{X}_k] \tag{2-2-25}$$

$$K_{k+1}=P_kh_{k+1}^T[h_{k+1}P_kh_{k+1}^T+\lambda]^{-1} \tag{2-2-26}$$

$$P_{k+1}=\frac{1}{\lambda}[1-K_{k+1}h_{k+1}]P_k \tag{2-2-27}$$

给出初值$\hat{X}_0$、P_0和遗忘因子λ，即可进行递推估计。

所以，对于噪声方差R_k为时变的情况，用最优加权最小二乘估计法得到的估计其方差最小，而用阻尼最小二乘估计法得到的方差增大，其方差与λ取值有关，故称为次优最小二乘估计法，用普通最小二乘估计法效果最差。

下面举几个最小二乘估计应用例子。

例2-2-2　设已知一脉冲雷达，其距离测量方差 $\sigma_R^2=2500\text{m}^2$，方位测量方差 $\sigma_\theta^2=0.0025\text{rad}^2$，对某一目标进行了六次测量。

六次测量数据如下：

i	1	2	3	4	5	6
$R_{(km)}$	10.10	10.08	9.90	9.94	10.06	9.98
$\theta°$	46.4°	44.6°	45.6°	46.0°	43.5°	44.5°

求目标的距离和方位估值。

解：因为距离和方位角测量是独立进行的，所以可分别求$\hat{R}$和$\hat{\theta}$

$$\hat{R}=\frac{1}{6}\sum_{i=1}^{6}R_i$$

$$=\frac{1}{6}(10.10+10.08+9.90+9.94+10.06+9.98)$$

$$=\frac{1}{6}\times 60.06$$

$$=10.01\text{km}$$

$$P_{\hat{R}}=\frac{1}{6}\sigma_R^2=\frac{1}{6}\times 2500=41.7\text{m}^2$$

$$\hat{\theta}=\frac{1}{6}\sum_{i=1}^{6}\theta_i$$

$$=\frac{1}{6}(46.4+44.6+45.6+46.0+43.5+44.5)$$

$$=\frac{1}{6}\times 270.6=45.1°$$

$$P_{\hat{\theta}}=\frac{1}{6}\sigma_\theta^2=\frac{1}{6}\times 0.0025=4\times 10^{-4}\text{rad}^2$$

可见，由于对多次测量进行了最小二乘估计，使得对目标定位质量大大提高。

例2-2-3　若有一等加速运动的目标，目标在任意时刻t的位置为

$$x(t)=x(0)+t\dot{x}(0)+\frac{t^2}{2}\ddot{x}(0)$$

今有一组位置观测值

$$z(kT)=x(kT)+e(kT)\qquad k=0,\ 1,\ \cdots,N。$$

其中$e(kT)$为白噪声序列，方差为σ^2，求初始值$x(0)$、$\dot{x}(0)$、$\ddot{x}(0)$的估计及任意时刻t时的距离。

解：列出观测方程

$$\boldsymbol{Z}=\boldsymbol{H}\boldsymbol{X}+\boldsymbol{e}$$

其中

$$\boldsymbol{Z}=\begin{pmatrix} z(0) \\ z(T) \\ \vdots \\ z(NT) \end{pmatrix} \quad \boldsymbol{X}=\begin{pmatrix} x(0) \\ \dot{x}(0) \\ \ddot{x}(0) \end{pmatrix} \quad \boldsymbol{e}=\begin{pmatrix} e(0) \\ e(T) \\ \vdots \\ e(NT) \end{pmatrix}$$

$$\boldsymbol{H}=\begin{pmatrix} 1 & 0 & 0 \\ 1 & T & T^2/2 \\ \cdots & \cdots & \cdots \\ 1 & NT & (NT)^2/2 \end{pmatrix}$$

$$E[\boldsymbol{e}]=0,\quad \boldsymbol{R}=E[\boldsymbol{e}\boldsymbol{e}^T]=\sigma^2 \boldsymbol{I}$$

估计矢量为

$$\hat{\boldsymbol{X}}=\begin{pmatrix} \hat{x}(0) \\ \hat{\dot{x}}(0) \\ \hat{\ddot{x}}(0) \end{pmatrix}$$

由最优加权最小二乘估计公式求得

$$\hat{\boldsymbol{X}}_{lsw}=(\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{Z}$$
$$=(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{Z}$$
$$\boldsymbol{P}_{lsw}=\sigma^2(\boldsymbol{H}^T\boldsymbol{H})^{-1}$$

把已知的$\boldsymbol{H}$、$\boldsymbol{Z}$代入就可求出$\hat{x}(0)$、$\hat{\dot{x}}(0)$、$\hat{\ddot{x}}(0)$及相应的方差，并得出在任意时刻t时目标距离的估计为

$$\hat{x}(t)=\left[1,t,\frac{t^2}{2}\right]\begin{pmatrix} \hat{x}(0) \\ \hat{\dot{x}}(0) \\ \hat{\ddot{x}}(0) \end{pmatrix}$$
$$=\hat{x}(0)+t\,\hat{\dot{x}}(0)+\frac{t^2}{2}\hat{\ddot{x}}(0)$$

例2-2-4　设有通信系统，发出的信号为$s(t)=a_k\cos k\omega_d t+b_k\sin k\omega_d t$，发射时间为$0\leqslant t\leqslant(N-1)T$，$k$——正整数，$0\leqslant k\leqslant\frac{N}{2}$，其中$\omega_d=\frac{\omega_s}{N}$，$\omega_s=\frac{2\pi}{T}$——采样角频率。接收机得到的观测数据为$z(nT)=s(nT)+e(nT)$，$n=0,1,\cdots,N-1$，$e(nT)$为白噪声序列，方差为$\sigma^2$，求对$a_k$、$b_k$的估计。

解：写出观测方程

$$Z = HX + e$$

其中

$$X = \begin{bmatrix} a_k \\ b_k \end{bmatrix}, \quad Z = \begin{pmatrix} z(0) \\ z(T) \\ \vdots \\ z(NT-T) \end{pmatrix}, \quad e = \begin{pmatrix} e(0) \\ e(T) \\ \vdots \\ e(NT-T) \end{pmatrix}$$

$$H = \begin{pmatrix} 1 & 0 \\ \cos k\omega_d T & \sin k\omega_d T \\ \cdots\cdots & \cdots\cdots \\ \cos k\omega_d (NT-T) & \sin k\omega_d (NT-T) \end{pmatrix}$$

根据巴塞伐尔定理

$$\sum_{n=0}^{N-1} |f_n|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |F_k|^2$$

得

$$\sum_{n=0}^{N-1} \cos^2 nk\omega_d T = \sum_{n=1}^{N-1} \sin^2 nk\omega_d T = \frac{N}{2}$$

$$\sum_{n=0}^{N-1} \cos nk\omega_d T \cdot \sin nk\omega_d T = 0$$

容易求得

$$H^T H = \frac{N}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

于是求得估计

$$\hat{a}_k = \frac{2}{N} \sum_{n=0}^{N-1} z(nT) \cos nk\omega_d T$$

$$\hat{b}_k = \frac{2}{N} \sum_{n=0}^{N-1} z(nT) \sin nk\omega_d T$$

估计误差方差为

$$P_{\hat{a}_k} = P_{\hat{b}_k} = \frac{2\sigma^2}{N}$$

若发射信号为一组合成音

$$s(t) = \sum_{k=k_1}^{k_m} (a_k \cos k\omega_d T + b_k \sin k\omega_d T) \qquad 0 < k_1 < k_m < \frac{N}{2}$$

利用上述结果，对于k_1，k_2，…，k_m，求出估计为

$$\hat{a}_k = \frac{2}{N} \sum_{n=0}^{N-1} z(nT) \cos nk\omega_d T$$

$$\hat{b}_k = \frac{2}{N}\sum_{n=0}^{N-1} z(nT)\sin nk\omega_d T$$

我们注意到

$$\hat{a}_k - j\hat{b}_k = \frac{2}{N}\sum_{n=0}^{N-1} z(nT)e^{-jnk\omega_d T}$$

显见，$\hat{a}_k - j\hat{b}_k$ 正好是 $DFT\{z(nT)\}$ 的一个分量，所以，对 a_k、b_k 的估计正好是对接收信号序列 $z(0)$，$z(1)$，…，$z(nT-T)$ 的谱估计。

§2-3 非线性最小二乘估计

在前两节讨论中，限定测量方程是线性的，即测量值与被估计量之间成线性关系。但是实际工作中常常遇到测量值与被估计量之间成非线性关系，这时测量方程表达成

$$Z = h(X) + V \tag{2-3-1}$$

其中 $h(X)$ 是 X 的非线性函数。

把 Z 在 $\bar{X}_0$ 邻近按泰勒级数展开，取其一阶近似，并令 $\bar{Z}_0 = h(\bar{X}_0)$，$\bar{X}_0$ 为 X 的初始估计。

$$dZ = Z - \bar{Z}_0 \approx \left.\frac{\partial h}{\partial X}\right|_{X=\bar{X}_0}(X - \bar{X}_0) + V^*$$

$$\triangleq \left.\frac{\partial h}{\partial X}\right|_{X=\bar{X}} dX + V^* \tag{2-3-2}$$

其中 $dX = X - \bar{X}_0$，V^* 为原来的测量噪声加上线性化余项。在 dZ 与 dX 之间存在线性关系，从而可以应用上述最小二乘估计公式。

例2-3-1 利用角度测量求目标位置估计。从位于同一基线上的观测点 B_i，$i=1, 2, \cdots, n$，对目标 A 进行方位测量，测量值为 $z_i = \theta_i$，求目标 A 的位置估计 $(\hat{x}, \hat{y})$。

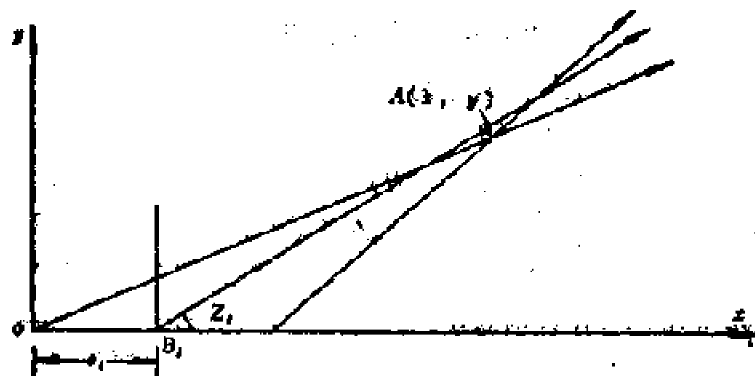


图 2-3-1 方位法对噪声源定位

解： z_i、l_i 与目标位置 (x, y) 之间存在下列关系

$$z_i = \tan^{-1}\frac{y}{x - l_i} + V_i \tag{2-3-3}$$

其中V_i为方位角测量随机误差，假设

$$E[V_i]=0,\qquad E[V_iV_j]=\begin{cases}\sigma_i^2 & i=j\\ 0 & i\neq j\end{cases}\tag{2-3-4}$$

若已给出目标位置的先验估计($\bar{x}$，$\bar{y}$)，写出增量方程，忽略线性化余项，得

$$z_i-\bar{z}_i=[H_{1i},H_{2i}]\begin{bmatrix}x-\bar{x}\\ y-\bar{y}\end{bmatrix}+V_i\tag{2-3-5}$$

其中
$$\bar{z}_i=\frac{\bar{y}}{\bar{x}-l_i}\tag{2-3-6}$$

$$H_{1i}=\left.\frac{\partial z_i}{\partial x}\right|_{\substack{x=\bar{x}\\ y=\bar{y}}}\qquad H_{2i}=\left.\frac{\partial z_i}{\partial y}\right|_{\substack{x=\bar{x}\\ y=\bar{y}}}\tag{2-3-7}$$

用矩阵符号表达，记

$$dZ=\begin{pmatrix}z_1-\bar{z}_1\\ z_2-\bar{z}_2\\ \vdots\\ z_n-\bar{z}_n\end{pmatrix}\qquad H=\begin{pmatrix}H_{11} & H_{21}\\ H_{12} & H_{22}\\ \cdots & \cdots\\ H_{1n} & H_{2n}\end{pmatrix}\tag{2-3-8}$$

$$dX=\begin{pmatrix}x-\bar{x}\\ y-\bar{y}\end{pmatrix}\qquad V=\begin{pmatrix}V_1\\ V_2\\ \vdots\\ V_n\end{pmatrix}$$

得出矢量观测方程为

$$dZ=H\,dX+V\tag{2-3-9}$$

已知
$$E[V]=0\qquad R=E[VV^T]=\begin{pmatrix}\sigma_1^2 & & 0\\ & \sigma_2^2 & \\ & & \ddots & \\ 0 & & & \sigma_n^2\end{pmatrix}\tag{2-3-10}$$

容易求出位置差 dX的估计误差协方差阵为

$$P=(H^TR^{-1}H)^{-1}\tag{2-3-11}$$

位置差dX的估计为
$$d\hat{X}=(H^TR^{-1}H)^{-1}H^TR^{-1}dZ\tag{2-3-12}$$

若 $d\hat{X}$显著大于零，我们应重复上述步骤，再围绕($\bar{x}+d\bar{x}$，$\bar{y}+d\bar{y}$)点展开。原则上，应重复以上步骤，直到求出的 $d\hat{x}\approx0$，$d\hat{y}=0$。协方差矩阵P决定了估计误差的不确定椭圆的大小与形状。

例 2-3-2　举一个数字例子，设$n=3$，给出数据如下：

B_i	1	2	3	单位
l_i	0	500	1000	m
z_i	30.1	45.0	73.6	deg
σ_i^2	0.01	0.01	0.04	deg²

开始时，给出位置的粗略估计，设为

$$\bar{x}=1210\text{m},\qquad \bar{y}=700\text{m}$$

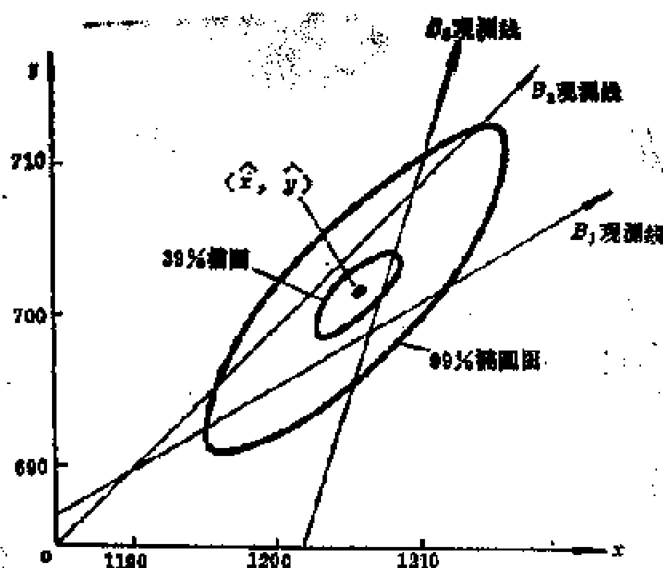


图 2-3-2 纯方位法定位的不确定椭圆

由此，可求出

$$dZ=\begin{pmatrix} .05 \\ .40 \\ .29 \end{pmatrix}\text{deg} \qquad H=\begin{pmatrix} -.0205 & .0354 \\ -.0403 & .0409 \\ -.0351 & .0225 \end{pmatrix}\text{deg}-\text{m}^{-1}$$

已知

$$R=\begin{pmatrix} .01 & & 0 \\ & .01 & \\ 0 & & .04 \end{pmatrix}\text{deg}^2$$

解出

$$P=(H^TR^{-1}H)^{-1}=\begin{pmatrix} 11.26 & 10.31 \\ 10.31 & 12.75 \end{pmatrix}\text{m}^2$$

$$d\hat{X}=PH^TR^{-1}dZ$$

$$=\begin{pmatrix} 13.4 & -3.1 & -15.3 \\ 24.0 & 10.6 & -12.2 \end{pmatrix}\cdot\begin{pmatrix} .05 \\ .40 \\ .29 \end{pmatrix}$$

$$=\begin{pmatrix} -5.01 \\ 1.89 \end{pmatrix}\text{m}$$

得 $\hat{x}=\overline{x}+\hat{d}x=1205.0\text{m}$

$\hat{y}=\overline{y}+\hat{d}y=701.9\text{m}$

如果我们对估计不满意，则重复上述步骤直到获得满意的结果为止。

求出 P的特征值和特征矢量如下：

特征值	特征矢量
22.34m²	43.0°
1.66m²	-43.0°

特征矢量方向就是不确定椭圆的长轴与短轴方向，以 $\sqrt{22.34}=4.72\text{m}$ 作为长轴半径，以$\sqrt{1.66}=1.29\text{m}$ 作为短轴半径，构成的椭圆为 39% 不确定椭圆，椭圆的圆心位于[1205.0，701.9]，把长短轴的半径扩大三倍构成的椭圆为99%不确定椭圆。

计算一下加权误差平方和有多大也是有意义的。

$$(\hat{z}_1 - z_1)^2 = (30.22 - 30.1)^2 = .0144\text{deg}^2$$
$$(\hat{z}_2 - z_2)^2 = (44.88 - 45.0)^2 = .0144\text{deg}^2$$
$$(\hat{z}_3 - z_2)^2 = (73.73 - 73.6)^2 = .0169\text{deg}^2$$

其中 $\hat{z}_i = \tan^{-1}\left(\dfrac{\hat{y}}{\hat{x} - l_i}\right)$

由此，得出加权误差平方和为

$$J(\hat{Z}) = \sum_{i=1}^{3} \frac{(\hat{z}_i - z_i)^2}{\sigma_i^2} = 1.44 + 1.44 + .42 = 3.30\text{deg}^2$$

若性能指标J不满足要求时，重复以上步骤，直到满意为止。这是一种迭代数值解法，常用来解非线性问题。

应当指出，非线性最小二乘解并不总是存在的，可能出现多值解，也可能没有解，而且与初值的选取有很大关系。总之，对非线性模型求最小二乘解目前尚存在困难，理论上有待进一步研究。

§2-4 最大似然估计$\hat{X}_{ml}$

在最小二乘估计中，我们并不要求知道观测值z_i的概率分布知识，现在假设已给出了观测值的概率密度函数形式，希望对其中的未知变量作出估计。

假定单次观测值z中包含未知参数x，x限定为确定量，我们可以写出z的条件密度函数$p(z/x)$，$p(z/x)$又称为似然函数，最大似然估计是指在$p(z/x)$达到极大时的x值，用符号$\hat{x}_{ml}$表示。得出$\hat{x}_{ml}$的必要条件是

$$\left.\frac{\partial p(z/x)}{\partial x}\right|_{x=\hat{x}_{ml}} = 0 \tag{2-4-1}$$

因为对数函数是单调函数，对$p(z/x)$取对数，不改变极大值的位置。为方便起见(2-4-1)式常用下式代替。

$$\left.\frac{\partial}{\partial x}\ln p(z/x)\right|_{x=\hat{x}_{ml}} = 0 \tag{2-4-2}$$

(2-4-1)式或(2-4-2)式称为极大似然估计方程。求解极大似然估计方程就得到$\hat{x}_{ml}$。

如果进行多次观测，令$Z^s = \{z_i\}$，$i = 1,2,\cdots,N$，表示观测值的集合，其中包含未知参数x，那么，条件概率密度函数为

$$p(Z^s/x) = p(z_1, z_2, \cdots, z_N/x)$$

若$z_1, z_2, \cdots, z_N$互相独立，则有

$$p(Z^s/x) = p(z_1/x)p(z_2/x)\cdots p(z_N/x)$$

在多次观测时，相应的最大似然估计方程为

$$\left.\frac{\partial p(Z^s/x)}{\partial x}\right|_{x=\hat{x}_{ml}} = 0 \tag{2-4-3}$$

或
$$\frac{\partial \ln p(Z^s/x)}{\partial x}\bigg|_{x=\hat{x}_{ml}} = 0 \tag{2-4-4}$$

例2-4-1 设有N个同高斯分布的独立观测值，试估计其均值和方差。

解：设观测值为z_i，$i=1,2,\cdots,N$，每个z_i的均值为μ，方差为σ^2，其条件密度函数为

$$p(z_i/\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[\frac{-(z_i-\mu)^2}{2\sigma^2}\right]$$

记观测集
$$Z^s=\{z_1,z_2,\cdots,z_N\}$$

则
$$p(Z^s/\mu,\sigma^2)=\prod_{i=1}^{N}p(z_i/\mu,\sigma^2)$$

$$=\frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left[-\sum_{i=1}^{N}(z_i-\mu)^2/2\sigma^2\right]$$

第一种情况，σ^2已知，μ未知。

由极大似然估计方程

$$\frac{\partial}{\partial\mu}\ln p(Z^s/\mu,\sigma^2)\bigg|_{\mu=\hat{\mu}_{ml}}=0 \tag{2-4-5}$$

容易解出
$$\hat{\mu}_{ml}=\frac{1}{N}\sum_{i=1}^{N}z_i \tag{2-4-6}$$

因为
$$E[\hat{\mu}_{ml}]=\frac{1}{N}\sum_{i=1}^{N}E[z_i]=\mu,$$

显见，$\hat{\mu}_{ml}$是μ的无偏估计。

第二种情况，μ已知，σ^2未知，由极大似然估计方程

$$\frac{\partial}{\partial\sigma^2}\ln p(Z^s/\mu,\sigma^2)\bigg|_{\sigma^2=\hat{\sigma}^2_{ml}}=0 \tag{2-4-7}$$

解出
$$\hat{\sigma}^2_{ml}=\frac{1}{N}\sum_{i=1}^{N}(z_i-\mu)^2 \tag{2-4-8}$$

因为 $E[\hat{\sigma}^2_m]=\sigma^2$，显见，这个估计量也是无偏的。

第三种情况，μ和σ^2均未知，由以上两个极大似然估计方程可求出

$$\hat{\mu}_{ml}=\frac{1}{N}\sum_{i=1}^{N}z_i \tag{2-4-9}$$

$$\hat{\sigma}^2_{ml}=\frac{1}{N}\sum_{i=1}^{N}(z_i-\hat{\mu}_{ml})^2 \tag{2-4-10}$$

其中的均值估计量是无偏的，但方差估计量不是无偏的。因为$z_i=\mu+e_i$

$$z_i-\hat{\mu}_{ml}=\mu+e_i-\mu-\frac{1}{N}\sum_{i=1}^{N}e_i=e_i-\frac{1}{N}\sum_{i=1}^{N}e_i$$

所以

$$E[\widehat{\sigma}^2_{ml}]=\frac{1}{N}\sum_{i=1}^{N}E\left[(e_i-\frac{1}{N}\sum_{i=1}^{N}e_i)^2\right]$$

$$=\frac{1}{N}\sum_{i=1}^{N}\left(\sigma_i^2-\frac{1}{N}\sigma_i^2\right)$$

$$=\frac{N-1}{N}\sigma^2$$

为了考察极大似然估计的效果，可以计算其偏量和方差，但在很多情况下，这样做往往很困难。于是，我们不是直接去处理这个问题，而是去推算出任意无偏估计的方差的下界，然后再考察 $\widehat{x}_{ml}$ 的方差和这个下界相比较的结果。

克拉美-罗方差下界

现在来研究变量 x 的任一无偏估计 $\widehat{x}$ 的方差，先证明下述定理。

若 $\widehat{x}$ 是 x 的任一无偏估计，则满足不等式

$$\mathrm{var}[\widehat{x}-x]\geqslant\left\{E\left[\frac{\partial}{\partial x}\ln p(z/x)\right]^2\right\}^{-1} \tag{2-4-11}$$

或等价于

$$\mathrm{var}[\widehat{x}-x]\geqslant-\left\{E\left[\frac{\partial^2}{\partial x^2}\ln p(z/x)\right]\right\}^{-1} \tag{2-4-12}$$

其中 (2-4-11)、(2-4-12)式的右端表达式称为**克拉美-罗方差下界**。意思是说，任一无偏估计的方差不可能小于克拉美罗下界。

(2-4-11)、(2-4-12)式的证明可利用施瓦兹不等式得出，由于假定 $\widehat{x}$ 是无偏的，故有

$$E[\widehat{x}-x]=\int_{-\infty}^{\infty}p(z/x)(\widehat{x}-x)dz=0 \tag{2-4-13}$$

两边对 x 求导，得

$$\frac{\partial}{\partial x}\int_{-\infty}^{\infty}(\widehat{x}-x)p(z/x)dz$$

$$=\int_{-\infty}^{\infty}\frac{\partial}{\partial x}\left[(\widehat{x}-x)p(z/x)dz\right]$$

$$=-\int_{-\infty}^{\infty}p(z/x)dz+\int_{-\infty}^{\infty}(\widehat{x}-x)\frac{\partial}{\partial x}p(z/x)dz$$

$$=0 \tag{2-4-14}$$

第一个积分等于 1 。注意到

$$\frac{\partial p(z/x)}{\partial x}=\frac{\partial\ln p(z/x)}{\partial x}\cdot p(z/x) \tag{2-4-15}$$

把(2-4-15)代入(2-4-14)式中，得

$$\int_{-\infty}^{\infty}\frac{\partial\ln p(z/x)}{\partial x}\cdot p(z/x)\cdot(\widehat{x}-x)dz=1 \tag{2-4-16}$$

或改写成

$$\int_{-\infty}^{\infty}\left[\frac{\partial \ln p(z/x)}{\partial x}\sqrt{p(z/x)}\right]\left[(\hat{x}-x)\sqrt{p(z/x)}\right]dz=1 \qquad (2\text{-}4\text{-}17)$$

应用施瓦兹不等式(参见附录A)

$$\left[\int_{-\infty}^{\infty} f(x)g(x)dx\right]^2 \leqslant \int_{-\infty}^{\infty}|f(x)|^2dx \cdot \int_{-\infty}^{\infty}|g(x)|^2dx$$

得到

$$\left\{\int_{-\infty}^{\infty}\left[\frac{\partial p(z/x)}{\partial x}\right]^2 p(z/x)dz\right\}\cdot\left\{\int_{-\infty}^{\infty}(\hat{x}-x)^2p(z/x)dz\right\}\geqslant 1 \qquad (2\text{-}4\text{-}18)$$

由施瓦兹不等式的推导可知，当对一切 z 和 x，满足下列条件时

$$\frac{\partial p(z/x)}{\partial x}=(\hat{x}-x)k(x) \qquad (2\text{-}4\text{-}19)$$

(2-4-18)式才能取等号，在 $k(x)$ 中不能有 $\hat{x}$。

由于 $$\int_{-\infty}^{\infty}(\hat{x}-x)^2p(z/x)dz=E[(\hat{x}-x)^2]$$

$$\int_{-\infty}^{\infty}\left[\frac{\partial \ln p(z/x)}{\partial x}\right]^2 p(z/x)dz=E\left\{\left[\frac{\partial \ln p(z/x)}{\partial x}\right]^2\right\}$$

代入(2-4-18)式中，得

$$E[(\hat{x}-x)^2]\geqslant\left\{E\left[\frac{\partial \ln p(z/x)}{\partial x}\right]^2\right\}^{-1} \qquad (2\text{-}4\text{-}20)$$

这就证明了(2-4-11)式。再来证明(2-4-12)式。

注意到 $$\int_{-\infty}^{\infty}p(z/x)dz=1$$

对 x 求导，得

$$\int_{-\infty}^{\infty}\frac{\partial p(z/x)}{\partial x}dz=\int_{-\infty}^{\infty}\frac{\partial \ln p(z/x)}{\partial x}\cdot p(z/x)dz=0 \qquad (2\text{-}4\text{-}21)$$

再对 x 求导一次，得

$$\int_{-\infty}^{\infty}\frac{\partial^2 \ln p(z/x)}{\partial x^2}p(z/x)dz+\int_{-\infty}^{\infty}\left[\frac{\partial \ln p(z/x)}{\partial x}\right]^2 p(z/x)dz=0 \qquad (2\text{-}4\text{-}22)$$

于是，可得出

$$E\left[\frac{\partial^2 \ln p(z/x)}{\partial x}\right]=-E\left[\frac{\partial \ln p(z/x)}{\partial x}\right]^2 \qquad (2\text{-}4\text{-}23)$$

代入(2-4-20)式，即得出

$$\mathrm{var}[\hat{x}-x]\geqslant-\left\{E\left[\frac{\partial^2 \ln p(z/x)}{\partial x}\right]\right\}^{-1} \qquad (2\text{-}4\text{-}24)$$

对以上结果可得出几个重要结论：

1. 对于任一无偏估计必然存在大于或等于某个界限的方差，这个界限就是克拉美-罗下界。

2. 若满足等式$\frac{\partial \ln p(z/x)}{\partial x}=(\hat{x}-x)K(x)$ (2-4-25)

则估计误差方差达到克拉美-罗下界。

3. 为了应用此下界，必须先验证估计量是无偏的。

4. 若观测样本数为N，且互相独立，则克拉美-罗下界为

$$\operatorname{var}[\hat{x}]=\frac{1}{NE\left\{\left[\frac{\partial}{\partial x}\ln p(z/x)\right]^2\right\}} \tag{2-4-26}$$

5. 若等式$\frac{\partial p(z/x)}{\partial x}=(\hat{x}-x)K(x)$不成立，将不知道这个估计的方差究竟是多大，但知道它一定大于克拉美-罗下限。

6. 在满足该式条件下，最大似然估计$\hat{x}_{ml}$是方差达到下界的无偏估计。证明如下：

假定 $$\left.\frac{\partial \ln p(z/x)}{\partial x}\right|_{x=\hat{x}_{ml}}=\left.(\hat{x}-x)K(x)\right|_{x=\hat{x}_{ml}}$$

上式左边为最大似然估计方程，必等于零。所以，为满足等式

$$\left.(\hat{x}-x)K(x)\right|_{x=\hat{x}_{ml}}=0$$

必定有$\hat{x}=\hat{x}_{ml}$。

换言之，在满足(2-4-25)式条件下，最大似然估计是方差达到下界的一个无偏估计。

文献上把方差达到克拉美-罗下界的估计称为优效估计。

例2-4-2 设有N个同为高斯分布的独立观测值，试求其均值的最大似然估计$\hat{\mu}_{ml}$，并求$\operatorname{var}[\hat{\mu}_{ml}]$。

解：已知 $$p(z/x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right]$$

和 $$p(Z^t/\mu)=\prod_{i=1}^{N}\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(z_i-\mu)^2}{2\sigma^2}\right]$$

我们有 $$\frac{\partial \ln p(Z^t/\mu)}{\partial \mu}=\frac{N}{\sigma^2}\left(\frac{1}{N}\sum_{i=1}^{N}z_i-\mu\right)$$

于是得 $$\hat{\mu}_{ml}=\frac{1}{N}\sum_{i=1}^{N}z_i$$

两边取统计平均

$$E[\hat{\mu}_{ml}]=\frac{1}{N}\sum_{i=1}^{N}E[z_i]=\mu$$

所以$\hat{\mu}_{ml}$是μ的无偏估计。

由于 $$\frac{\partial \ln p(Z^t/\mu)}{\partial \mu}=\frac{N}{\sigma^2}(\hat{\mu}_{ml}-\mu)$$

符合(2-4-25)式要求，故$\hat{\mu}_{ml}$为优效估计。于是得出估计误差方差阵为

$$P=-\left(\frac{\partial^2\ln p(Z/X)}{\partial X^2}\right)^{-1}$$

一般地说，当测量Z为非高斯分布时，最优加权最小二乘估计与最大似然估计不一致。而当Z为高斯分布时，这二种估计相同。

多参数最大似然估计的方差下界

直接求$\hat{X}_{ml}$的方差阵，除了特殊情况外，一般说来都是很困难的。类似于标量最大似然估计的克拉美-罗方差下界，对于矢量最大似然估计，亦存在矢量克拉美-罗方差下界。

记

$$J=E\left[\left(\frac{\partial\ln p(Z/X)}{\partial X}\right)\left(\frac{\partial\ln p(Z/X)}{\partial X}\right)^T\right]$$

$$=-E\left\{\frac{\partial}{\partial X}\left[\frac{\partial}{\partial X}\ln p(Z/X)\right]^T\right\} \tag{3-4-27}$$

矩阵J中的第ij元素(第i行第j列)为

$$J_{ij}=E\left[\frac{\partial\ln p(Z/X)}{\partial x_i}\cdot\frac{\partial\ln p(Z/X)}{\partial x_j}\right]$$

$$=-E\left[\frac{\partial^2\ln p(Z/X)}{\partial x_i\partial x_j}\right] \tag{2-4-28}$$

矩阵J称为**费什尔信息矩阵**。

费什尔信息矩阵的逆阵称为关于矢量X的克拉美-罗方差矩阵下界，记作

$$(CRMB)_X=J^{-1}$$

J^{-1}的第ij元素记作J^{ij}。则存在以下结果：

对于元素x_i的任意无偏估计$\hat{x}_i$，其估计误差满足不等式

$$\sigma_i^2=E[(\hat{x}_i-x)^2]\geqslant J^{ii} \tag{2-4-29}$$

上式中取等号的条件是对于所有x_i均成立下列关系

$$\hat{x}_i-x_i=\sum_{j=1}^{k}K_{ij}(X)\frac{\partial\ln p(Z/X)}{\partial x_j} \tag{2-4-30}$$

其中$K_{ij}(X)$是与Z无关，但可能依赖于X的任意常数，而$\hat{x}_i$则是只依赖于Z的任意函数。换言之，当估计误差可以表达为各偏导数$\frac{\partial\ln p(Z/X)}{\partial x_i}$的加权和时，$\hat{x}_i$成为优效估计，其方差达到克拉美-罗下界$J^{ii}$。

现在证明$i=1$时，(2-4-29)式成立。

令

$$F=\begin{pmatrix}\hat{x}_1-x_1\\ \frac{\partial\ln p(Z/X)}{\partial x_1}\\ \vdots\\ \frac{\partial\ln p(Z/X)}{\partial x_k}\end{pmatrix} \tag{2-4-31}$$

其协方差阵为

$$Q=E[FF^T]=\begin{pmatrix}\sigma_1^2 & 1 & 0 & \cdots & 0\\ 1 & J_{11} & J_{12} & \cdots & J_{1k}\\ 0 & J_{21} & J_{22} & \cdots & J_{2k}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & J_{k1} & J_{k2} & \cdots & J_{kk}\end{pmatrix} \tag{2-4-32}$$

$$=\begin{pmatrix} E[(\hat{x}_1-x_1)^2] & E\left[(\hat{x}_1-x_1)\frac{\partial \ln p(Z/X)}{\partial x_1}\right] & \cdots & E\left[(\hat{x}_1-x_1)\frac{\partial \ln p(Z/X)}{\partial x_k}\right]\\ E\left[\frac{\partial \ln p(Z/X)}{\partial x_1}(\hat{x}_1-x_1)\right] & E\left[\left(\frac{\partial \ln p(Z/X)}{\partial x_1}\right)^2\right] & \cdots & E\left[\frac{\partial \ln p(Z/X)}{\partial x_1}\cdot\frac{\partial \ln p(Z/X)}{\partial x_k}\right]\\ \cdots & \cdots & \cdots & \cdots\\ E\left[\frac{\partial \ln p(Z/X)}{\partial x_k}(\hat{x}_1-x_1)\right] & \cdots & \cdots & E\left[\left(\frac{\partial \ln p(Z/X)}{\partial x_k}\right)^2\right]\end{pmatrix}$$

由于Q是一个协方差矩阵，故必为非负定，其行列式必大于或等于零，即有

$$|Q|\geqslant 0 \tag{2-4-33}$$

因为行列式

$$|Q|=\sigma_1^2\begin{vmatrix}J_{11} & J_{12} & \cdots & J_{1k}\\ J_{21} & J_{22} & \cdots & J_{2k}\\ \cdots & \cdots & \cdots & \cdots\\ J_{k1} & J_{k2} & \cdots & J_{kk}\end{vmatrix}-\begin{vmatrix}1 & 0 & \cdots & 0\\ J_{21} & J_{22} & \cdots & J_{2k}\\ \cdots & \cdots & \cdots & \cdots\\ J_{k1} & J_{k2} & \cdots & J_{kk}\end{vmatrix} \tag{2-4-34}$$

其中

$$\begin{vmatrix}J_{11} & J_{12} & \cdots & J_{1k}\\ J_{21} & J_{22} & \cdots & J_{2k}\\ \cdots & \cdots & \cdots & \cdots\\ J_{k1} & J_{k2} & \cdots & J_{kk}\end{vmatrix}=|J| \tag{2-4-35}$$

$$\begin{vmatrix}1 & 0 & \cdots & 0\\ J_{21} & J_{22} & \cdots & J_{kk}\\ \cdots & \cdots & \cdots & \cdots\\ J_{k1} & J_{k2} & \cdots & J_{kk}\end{vmatrix}=J_{11}\text{的代数余子式} \tag{2-4-36}$$

于是得到

$$\sigma_1^2|J|-J_{11}\text{的代数余子式}\geqslant 0 \tag{2-4-37}$$

或

$$\sigma_1^2\geqslant\frac{J_{11}\text{的代数余子式}}{|J|}=J^{11} \tag{2-4-38}$$

这就证明了$i=1$时(2-4-29)式成立，同理，可以证明$i=2,3,\cdots,k$时(2-4-29)式亦成立。

再来证明(2-4-30)式，为使(2-4-29)式中的$\geqslant$号成为等号，必需也只需$|Q|=0$，根据行列式理论得知，一个行列式等于零的充分必要条件是它的各列（或各行）向量相互线性依赖。因此，为使$|Q|=0$，必需也只需其第一列元素能够表示成其余各列元素的线性组合，即存在一组常数$K_{1j}(x)$，$j=1,2,\cdots,k$，使得下面诸式成立：

$$E[(\hat{x}_1-x_1)^2]=\sum_{j=1}^{k}K_{1j}(X)E\left[(\hat{x}_1-x_1)\frac{\partial\ln p(Z/X)}{\partial x_j}\right]$$

$$=E\left\{(\hat{x}_1-x_1)\left[\sum_{j=1}^{k}K_{1j}(X)\frac{\partial\ln p(Z/X)}{\partial x_j}\right]\right\} \tag{2-4-39}$$

$$E\left[\frac{\partial\ln p(Z/X)}{\partial x_1}(\hat{x}_1-x_1)\right]$$

$$=\sum_{j=1}^{k}K_{1j}(X)E\left[\frac{\partial\ln p(Z/X)}{\partial x_1}\cdot\frac{\partial\ln p(Z/X)}{\partial x_j}\right]$$

$$=E\left\{\frac{\partial\ln p(Z/X)}{\partial x_1}\left[\sum_{j=1}^{k}K_{1j}(X)\frac{\partial\ln p(Z/X)}{\partial x_j}\right]\right\} \tag{2-4-40}$$

…………………………

$$E\left[\frac{\partial\ln p(Z/X)}{\partial x_k}(\hat{x}_1-x_1)\right]$$

$$=\sum_{j=1}^{k}K_{1j}(X)E\left[\frac{\partial\ln p(Z/X)}{\partial x_k}\cdot\frac{\partial\ln p(Z/X)}{\partial x_j}\right]$$

$$=E\left\{\frac{\partial\ln p(Z/X)}{\partial x_k}\left[\sum_{j=1}^{k}K_{1j}(X)\frac{\partial\ln p(Z/X)}{\partial x_j}\right]\right\} \tag{2-4-41}$$

为使以上诸式对一切Z和X均成立，要求$(\hat{x}_1-x_1)$能表达成

$$\hat{x}_1-x_1=\sum_{j=1}^{k}K_{1j}(X)\frac{\partial\ln p(Z/X)}{\partial x_j} \tag{2-4-42}$$

这就证明了(2-4-30)式在$i=1$时成立，同理可证明$i=2,3,\cdots,k$时(2-4-30)式亦成立。

若联合估计$\hat{X}=(\hat{x}_1,\hat{x}_2,\cdots,x_k)^T$为无偏且其每一个分量的方差都达到各自的下界$J^{ii}$，则称估计$\hat{X}(Z)$达到了克拉美-罗下界，$\hat{X}(Z)$称为优效估计。(2-4-30)式称为优效估计条件。

(2-4-30)式还可写成矢量形式，得

$$\hat{X}-X=K\left[\frac{\partial}{\partial X}\ln p(Z/X)\right] \tag{2-4-43}$$

或
$$\frac{\partial}{\partial X}\ln p(Z/X)=K^{-1}(\hat{X}-X)=H(\hat{X}-X) \tag{2-4-44}$$

其中H为与Z无关的任意方阵。

非线性最大似然估计

设测量方程是非线性的，

$$z_i=h(x)+e_i \qquad i=1,2,\cdots,N$$

其中e_i为零均值高斯噪声，方差为σ^2，互相独立，求x的最大似然估计$\hat{X}_{ml}$。

由最大似然估计方程得

$$\left.\frac{\partial\ln p(z^N/x)}{\partial x}\right|_{x=\hat{x}_{ml}}=\frac{1}{\sigma^2}\sum_{i=1}^{N}\left[z_i-h(x)\right]\left.\frac{\partial h(x)}{\partial x}\right|_{x=\hat{x}_{ml}}=0$$

或
$$\frac{1}{\sigma^2}\frac{\partial h(x)}{\partial x}\left[\frac{1}{N}\sum_{i=1}^{N} z_i - h(x)\right]\Bigg|_{x=\hat{x}_{ml}} = 0 \tag{2-4-45}$$
解出
$$h(\hat{x}_{ml}) = \frac{1}{N}\sum_{i=1}^{N} z_i \tag{2-4-46}$$
或
$$\hat{x}_{ml} = h^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} z_i\right) \tag{2-4-47}$$
其中 h^{-1} 为反函数。

注意(2-4-45)式一般不满足优效估计条件，因而不能求出其方差，但可求出方差的下界。对(2-4-45)式再求导一次，得
$$\frac{\partial^2 \ln p(z'/x)}{\partial x^2} = \frac{1}{\sigma^2}\sum_{i=1}^{N}[z_i - h(x)]\frac{\partial^2 h(x)}{\partial x^2}$$
$$-\frac{N}{\sigma^2}\left[\frac{\partial h(x)}{\partial x}\right]^2 \tag{2-4-48}$$
两边取统计平均，由于$E[z_i - h(x)] = 0$，上式成为
$$E\left[\frac{\partial^2 \ln p(z'/x)}{\partial x^2}\right] = -\frac{N}{\sigma^2}E\left[\frac{\partial h(x)}{\partial x}\right]^2$$
$$= -\frac{N}{\sigma^2}\left[\frac{\partial h(x)}{\partial x}\right] \tag{2-4-49}$$
把以上结果代入克拉美-罗不等式，就得到
$$\mathrm{var}[\hat{x}_{ml} - x] \geq \frac{\sigma^2}{N\left[\left(\frac{\partial h(x)}{\partial x}\right)^2\right]_{x=\hat{x}_{ml}}} \tag{2-4-50}$$

§2-5 最小二乘拟合

数据拟合就是用一确定的数学表达式来近似表达某随机变量与其它确定变量之间的关系，这类确定变量可能是温度，时间等。在科学实验与工农业生产中有广泛应用。最小二乘法是数据拟合的常用方法，依据的准则是使拟合误差的平方和达到最小。所以，最小二乘拟合可以看成是最小二乘估计的一种应用。分下列几种类型进行讨论。

(一) 线性拟合

假设某一随机变量 y 是另一确定变量 x 的线性函数，x_i 是已知量，可表达为
$$y_i = a + bx_i + e_i \tag{2-5-1}$$
其中 e_i 为零均值白噪声，方差等于 σ^2。

例如，在硝酸银溶解度试验中，测得在不同温度下，溶解于100份水中的硝酸银份数 y 的数据如下表所示，用 x 代表温度，它是已知确定变量，而硝酸银的溶解度带有随机性，故 y 看成是随机变量。

x_i(°C)	0	4	10	15	21	29	36	51	68
y_i	66.7	71.0	76.3	80.6	85.7	92.9	99.4	113.6	125.1

将每对(x_i, y_i)在直角系统中画出，显见y_i与x_i近似成线性关系，因此用直线

$$\hat{y} = \hat{a} + \hat{b}x \tag{2-5-2}$$

进行拟合，现在的问题是要根据数组(x_i, y_i)来得到$\hat{a}$和$\hat{b}$的值。

采用最小二乘估计法，使用的准则是使误差平方和

$$J = \sum_{i=1}^{n}(y_i - \hat{a} - \hat{b}x_i)^2 \tag{2-5-3}$$

达到极小。

令下述偏导数等于零，

$$\frac{\partial J}{\partial \hat{a}} = -2\sum_{i=1}^{n}(y_i - \hat{a} - \hat{b}x_i)x_i = 0 \tag{2-5-4}$$

$$\frac{\partial J}{\partial \hat{b}} = -2\sum_{i=1}^{n}(y_i - \hat{a} - \hat{b}x_i)x_i = 0 \tag{2-5-5}$$

得到方程组

$$n\hat{a} + n\bar{x}\hat{b} = n\bar{y}$$

$$n\bar{x}\hat{a} + \sum_{i=1}^{n}x_i^2\hat{b} = \sum_{i=1}^{n}x_iy_i$$

其中 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$， $\bar{y} = \frac{1}{n}\sum_{i=1}^{n}y_i$

解出

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \tag{2-5-6}$$

$$\hat{b} = \frac{\sum_{i=1}^{n}(x_iy_i) - n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2 - n\bar{x}^2}$$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2-5-7}$$

把$\hat{a}$、$\hat{b}$代入(2-5-2)式中，得拟合方程为

$$\hat{y} = \hat{a} + \hat{b}x = \bar{y} + \hat{b}(x - \bar{x}) \tag{2-5-8}$$

对于上述已知数据组进行计算，得

x_i	y_i	x_i^2	y_i^2	x_iy_i
0	66.7	0	4448.9	0
4	71.0	16	5041.0	284.0
10	76.3	100	5821.7	763.0
15	80.6	225	6496.4	1209.0
21	85.7	441	7344.5	1799.7
29	92.9	841	8630.4	2694.1
36	99.4	1296	9880.4	3578.4
51	113.6	2601	12905.0	5793.6
68	125.1	4624	15650.0	8506.8
$\sum =$ 234	811.3	10144	762182	24628.6

$$\overline{x} = \frac{234}{9} = 26$$

$$\overline{y} = \frac{811.3}{9} = 90.14$$

$$\sum_{i=1}^{9}(x_i - \overline{x})^2 = 10144 - \frac{234^2}{9} = 4060$$

$$\sum_{i=1}^{9}(y_i - \overline{y})^2 = 76218.2 - \frac{[811.3}{9} = 3084$$

$$\sum_{i=1}^{9}(x_i - \overline{x})y_i = 24628.6 - \frac{234 \times 811.3}{9} = 3534.8$$

结果得

$$\hat{b} = \frac{3534.8}{4060} = 0.8706$$

$$\hat{a} = 67.5078$$

得出拟合线性方程为

$$\hat{y} = 67.5078 + 0.8706x$$

利用此方程式可进行内插或外推，但外推的置信度将大大下降。

若已知变量 x_i 是时间 t_i，随机变量 y_i 是一时间序列，则可得到线性拟合方程

$$\hat{y} = a + bt$$

上述方法容易推广到多元线性拟合，设随机变量是确定变量 x_1、x_2 的线性函数

$$y = a + b_1x_1 + b_2x_2 + e \qquad e \sim N(0, \sigma^2)$$

其中 a、b_1、b_2 及 σ^2 为不依赖于 x_1、x_2 的未知参数，现在根据样本集(x_{1i}, x_{2i}, y_i)，$i = 1, 2, \cdots, n$，利用最小二乘法求出二元线性拟合方程

$$\hat{y} = \hat{a} + \hat{b}_1x_1 + \hat{b}_2x_2 \qquad (2\text{-}5\text{-}9)$$

估计误差的平方和为

$$J=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

$$=\sum_{i=1}^{n}(y_i-\hat{a}-\hat{b}_1x_{1i}-\hat{b}_2x_{2i})^2 \qquad (2\text{-}5\text{-}10)$$

令 J 对 $\hat{a}$、$\hat{b}_1$、$\hat{b}_2$ 的偏导数等于零，得

$$\left.\begin{aligned}
&n\hat{a}+\hat{b}_1\sum_{i=1}^{n}x_{1i}+\hat{b}_2\sum_{i=1}^{n}x_{2i}=\sum_{i=1}^{n}y_i\\
&\hat{a}\sum_{i=1}^{n}x_{1i}+\hat{b}_1\sum_{i=1}^{n}x_{1i}^2+\hat{b}_2\sum_{i=1}^{n}x_{1i}x_{2i}=\sum_{i=1}^{n}x_{1i}y_i\\
&\hat{a}\sum_{i=1}^{n}x_{2i}+\hat{b}_1\sum_{i=1}^{n}x_{1i}x_{2i}+\hat{b}_2\sum_{i=1}^{n}x_{2i}^2=\sum_{i=1}^{n}x_{2i}y_i
\end{aligned}\right\} \qquad (2\text{-}5\text{-}11)$$

解得

$$\left.\begin{aligned}
&\hat{a}=\overline{y}-\hat{b}_1\overline{x}_1-\hat{b}_2\overline{x}_2\\
&\hat{b}_1\sum_{i=1}^{n}x_{1i}(x_{1i}-\overline{x}_1)+\hat{b}_2\sum_{i=1}^{n}(x_{2i}-\overline{x}_2)=\sum_{i=1}^{n}x_{1i}(y_i-\overline{y})\\
&\hat{b}_1\sum_{i=1}^{n}x_{2i}(x_{1i}-\overline{x}_1)+\hat{b}_2\sum_{i=1}^{n}x_{2i}(x_{2i}-\overline{x}_2)=\sum_{i=1}^{n}x_{2i}(y_i-\overline{y})
\end{aligned}\right\} \qquad (2\text{-}5\text{-}12)$$

引入记号

$$S(\tilde{x}_1^2)=\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)^2$$

$$S(\tilde{x}_2^2)=\sum_{i=1}^{n}(x_{2i}-\overline{x}_2)^2$$

$$S(\tilde{x}_1,\tilde{x}_2)=\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)(x_{2i}-\overline{x}_2)$$

$$S(\tilde{x}_1,\tilde{y})=\sum_{i=1}^{n}(x_{1i}-\overline{x}_1)(y_i-\overline{y})$$

$$S(\tilde{x}_2,\tilde{y})=\sum_{i=1}^{n}(x_{2i}-\overline{x}_2)(y_i-\overline{y})$$

把(2-5-12)式改写成

$$\left.\begin{aligned}
&\hat{b}_1S(\tilde{x}_1^2)+\hat{b}_2S(\tilde{x}_1,\tilde{x}_2)=S(\tilde{x}_1,\tilde{y})\\
&\hat{b}_1S(\tilde{x}_1,\tilde{x}_2)+\hat{b}_2S(\tilde{x}_2^2)=S(\tilde{x}_2,\tilde{y})
\end{aligned}\right\} \qquad (2\text{-}5\text{-}13)$$

若(2-5-13)式中系数行列式不为零，则可解出

$$\left.\begin{aligned}\hat{b}_1 &= \frac{1}{\Delta}[S(\tilde{x}_2^2)S(\tilde{x}_1, \tilde{y}) - S(\tilde{x}_1, \tilde{x}_2)S(\tilde{x}_2, \tilde{y})]\\ \hat{b}_2 &= \frac{1}{\Delta}[S(\tilde{x}_1^2)S(\tilde{x}_1, \tilde{y}) - S(\tilde{x}_1, \tilde{x}_2)S(\tilde{x}_1, \tilde{y})]\\ \hat{a} &= \bar{y} - \hat{b}_1\bar{x}_1 - \hat{b}_2\bar{x}_2\end{aligned}\right\} \quad (2\text{-}5\text{-}14)$$

及

其中 $\Delta = S(\tilde{x}_1^2)S(\tilde{x}_2^2) - (S(\tilde{x}_1, \tilde{x}_2))^2$。

由此可得二元线性拟合方程式为

$$\hat{y} = \bar{y} + \hat{b}_1(x_1 - \bar{x}_1) + \hat{b}_2(x_2 - \bar{x}_2) \quad (2\text{-}5\text{-}15)$$

这是三维空间中的一根直线。

（二） 多项式拟合

把上述线性拟合加以推广，用下列m次多项式

$$p_m(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m$$

$$y = \sum_{i=0}^{m} a_i x^i \quad (2\text{-}5\text{-}16)$$

来拟合一组数据(x_i, y_i)，$i = 0, 1, \cdots, n$。

仍然采用最小二乘法，使估计误差平方和达到最小。估计误差平方和为

$$\begin{aligned}J &= \sum_{i=0}^{n}[p_m(x_i) - y_i]^2\\ &= \sum_{i=0}^{n}\left[\sum_{j=0}^{m} a_j x_i^j - y_i\right]^2 \quad (2\text{-}5\text{-}17)\end{aligned}$$

令 J对$a_0, a_1, \cdots, a_m$的偏导数为零，得

$$\frac{\partial J}{\partial a_k} = \sum_{i=0}^{n} 2[p_m(x_i) - y_i]\frac{\partial p_m(x_i)}{\partial a_k} = 0 \qquad k = 0, 1, \cdots, m \quad (2\text{-}5\text{-}18)$$

改写(2-5-18)式，并引入符号

$$S_l = \sum_{i=0}^{n} x_i^l \quad (2\text{-}5\text{-}19)$$

$$t_k = \sum_{i=0}^{n} y_i x_i^k \quad (2\text{-}5\text{-}20)$$

令

$$\begin{aligned}\frac{\partial J}{\partial a_k} &= 2\sum_{i=0}^{n}[p_m(x_i) - y_i]x_i^k\\ &= 2\sum_{i=0}^{n}\left[\sum_{j=0}^{m} a_j x_i^{j+k} - y_i x_i^k\right]\\ &= 2\left[\sum_{j=0}^{m} a_j \sum_{i=0}^{n} x_i^{j+k} - \sum_{i=0}^{n} y_i x_i^k\right]\\ &= 2\left[\sum_{j=0}^{m} a_j S_{j+k} - t_k\right] = 0 \quad (2\text{-}5\text{-}21)\end{aligned}$$

注意 $$S_0=\sum_{i=0}^{n}x_i^0=1+1+\cdots+1=n+1$$

把(2-5-21)式展开，得方程组

$$\left.\begin{aligned}S_0a_0+S_1a_1+\cdots+S_ma_m=t_0\\S_1a_0+S_2a_1+\cdots+S_{m+1}a_m=t_1\\\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\S_ma_0+S_{m+1}a_1+\cdots+S_{2m}a_m=t_m\end{aligned}\right\}\qquad(2\text{-}5\text{-}22)$$

由(2-5-22)式可解出多项式$p(m)$的各系数，(2-5-22)式常称为正规方程。

例 已知一组实验数据

i	0	1	2	3	4
x_i	-2	-1	0	1	2
y_i	35.2	35.9	36.7	37.4	38.4

试用二次多项式

$$p_2(x)=a_0+a_1x+a_2x^2$$

进行最小二乘拟合。

解：本题的正规方程为

$$S_0a_0+S_1a_1+S_2a_2=t_0$$
$$S_1a_0+S_2a_1+S_3a_2=t_1$$
$$S_2a_0+S_3a_1+S_4a_2=t_2$$

由题意可算出 $S_0=5$，　$S_1=0$，　$S_2=10$，　$S_3=0$，　$S_4=34$，

$t_0=183.6$，　$t_1=7.9$，　$t_2=367.7$

解出 $\hat{a}_0=36.649$，　$\hat{a}_1=0.79$，　$\hat{a}_2=0.0357$

故得 $y=p_2(x)=36.649+0.79x+0.0357x^2$

(三) 函数拟合

最小二乘拟合方法还可推广到应用诸如指数函数、对数函数、三角函数等拟合观测数据。

例如用$y=ae^{bx}$去拟合实验数据集(x_i,y_i)，$i=0,1,\cdots,n$，其中a、b为未知常数，估计误差的平方和为

$$J=\sum_{i=0}^{n}(ae^{bx_i}-y_i)^2$$

令J对a、b的偏导数等于零，得

$$\frac{\partial J}{\partial a}=2\sum_{i=0}^{n}(ae^{bx_i}-y_i)e^{bx_i}=0$$

$$\frac{\partial J}{\partial b}=2a\sum_{i=0}^{n}(ae^{bx_i}-y_i)x_ie^{bx_i}=0$$

由于方程组为超越方程，不易求得解$\hat{a}$、$\hat{b}$。为此对原拟合方程作变换，

$$\lg y = \lg(a e^{bx}) = \lg a + b x \lg e$$

令 $u = \lg y$, $a_0 = \lg a$, $a_1 = b \lg e$

则有
$$u = a_0 + a_1 x$$

故拟合方程变成一元线性方程，于是对数据集(x_i, u_i)，$i = 0, 1, 2, \cdots, N$，作线性拟合，算出a_0和a_1，然后由

$$a = 10^{a_0}, \qquad b = \frac{a_1}{\lg e}$$

求出$\hat{a}$和$\hat{b}$，于是得拟合方程式

$$y = \hat{a} e^{\hat{b}x} = 10^{a_0} e^{\frac{a_1}{\lg e}x}$$

§2-6 有约束最小二乘估计

如果有测量方程

$$\underset{m\times 1}{Z} = \underset{m\times n}{H}\ \underset{n\times 1}{X} + \underset{m\times 1}{e} \qquad m > n \tag{2-6-1}$$

已知$E[e] = 0$，$E[ee^T] = R$

那末根据估计误差平方和

$$J = (Z - H\hat{X})^T(Z - HX) = \| (Z - H\hat{X}) \|^2 \tag{2-6-2}$$

达到最小的准则，可以得到X的最小二乘估计

$$\hat{X}_{ls} = (H^T H)^{-1} H^T Z \tag{2-6-3}$$

$$P_{ls} = (H^T H)^{-1} H^T R H (H^T H)^{-1} \tag{2-6-4}$$

上述$\hat{X}_{ls}$称为无约束最小二乘估计。在许多情况下，需要对$\hat{X}$加上某种约束条件，在约束条件下的最小二乘估计称为有约束最小二乘估计。

例如，有一种约束条件是规定测量信号能量与理想信号能量相同，这时约束条件表为达

$$X^T X = Z^T Z = c \tag{2-6-5}$$

用拉格朗日乘数法求约束条件下的估计。给出性能指标函数

$$J(\hat{X}, \lambda) = (Z - H\hat{X})^T(Z - H\hat{X}) + \lambda(\hat{X}^T\hat{X} - c) \tag{2-6-6}$$

对$\hat{X}$求偏导数，并使之等于零

$$\frac{\partial}{\partial \hat{X}} J(\hat{X}, \lambda) = 0 \tag{2-6-7}$$

$$-2H^T Z + 2H^T H\hat{X} + 2\lambda\hat{X} = 0$$

解出
$$\hat{X} = (H^T H + \lambda I)^{-1} H^T Z \tag{2-6-8}$$

再选择λ，使其满足约束条件(2-6-5)式

由
$$[(H^T H + \lambda I)^{-1} H^T Z]^T[(H^T H + \lambda I)^{-1} H^T Z] = c \tag{2-6-9}$$

解出$\lambda = \lambda_{opt}$代回(2-6-8)式，就得到有约束最小二乘估计

$$\hat{X}=(H^TH+\lambda_{opt}I)^{-1}H^TZ \tag{2-6-10}$$

另一个约束条件的例子，设Q为X的某个线性算子，要求在约束条件

$$\|Z-H\hat{X}\|^2=\|e\|^2=c \tag{2-6-11}$$

之下使$\|Q\hat{X}\|^2$达到最小。$\|X\|$表示矢量X的范数。

应用拉格朗日乘数法构造性能指标函数

$$J(\hat{X},\lambda)=\|Q\hat{X}\|^2+\lambda(\|Z-H\hat{X}\|-c) \tag{2-6-12}$$

对$\hat{X}$求偏导数，并使之等于零

$$\frac{\partial}{\partial\hat{X}}J(\hat{X},\lambda)=0 \tag{2-6-13}$$

$$2Q^TQ\hat{X}-2\lambda H^TZ+2\lambda H^TH\hat{X}=0$$

得

$$\hat{X}=(H^TH+\frac{1}{\lambda}Q^TQ)H^TZ \tag{2-6-14}$$

显然估计$\hat{X}$与变换矩阵Q有关，例如我们希望矢量$\hat{X}$的各分量比较平滑，这时可把$\hat{X}$的二阶差分的能量压到最小。设

$$X=[x_1,x_2,\cdots,x_i,\cdots,x_n]^T$$

则x_i的一阶差分　　$\nabla x_i=x_{i-1}-x_i$

x_i的二阶差分　　$\nabla^2x_i=x_{i-1}-2x_i+x_{i+1}$

于是有

$$\begin{pmatrix}\nabla x_1\\ \nabla^2x_2\\ \vdots\\ \vdots\\ \nabla^2x_n\end{pmatrix}=\begin{pmatrix}-2&1&&&&\\ 1&-2&1&&0&\\ &1&-2&1&&\\ &&\ddots&\ddots&\ddots&\\ 0&&&1&-2&1\\ &&&&1&-2\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ \vdots\\ \vdots\\ x_n\end{pmatrix} \tag{2-6-15}$$

或写成

$$\nabla^2X=QX \tag{2-6-16}$$

这里Q是一平滑化矩阵，它使信号矢量得到某种程度的平滑。

现在还要调节拉格朗日乘子λ，使得$\hat{X}$满足约束条件

$$\|Z-H\hat{X}\|^2=\|e\|^2=c \tag{2-6-17}$$

定义残差

$$\begin{aligned}\Gamma&=Z-H\hat{X}\\ &=Z-H(H^TH+\frac{1}{\lambda}Q^TQ)^{-1}H^TZ\end{aligned} \tag{2-6-18}$$

如果约束条件(2-6-17)式被满足，则表明$\hat{X}=X$，$\|\Gamma\|^2=c$，实际上只能做到

$$\|\Gamma\|^2=c\pm a \tag{2-6-19}$$

α 是一小量，称为准确度因子。

可用迭代法找出满足(2-6-19)式的 λ。第一步指定一个初值 λ_0，第二步按(2-6-14)式计算 $\hat{X}$ 和按(2-6-18)式计算 $\|\Gamma\|^2$，若(2-6-19)式被满足，则停止，否则，改变 λ_0，重新计算 $\hat{X}$ 和 $\|\Gamma\|^2$，直到(2-6-19)式满足为止。第三步，把找到的 λ_{opt} 值代入(2-6-14)式，就得到有约束最小二乘估计

$$\hat{X} = (H^T H + \frac{1}{\lambda_{opt}} Q^T Q) H^T Z \tag{2-6-20}$$

因为 $\|\Gamma\|^2 = \left[Z - H\left(H^T H + \frac{1}{\lambda} Q^T Q\right)^{-1} H^T Z\right]^T \left[Z - H\left(H^T H + \frac{1}{\lambda} Q^T Q\right)^{-1} H^T Z\right]$ 是 λ 的单调函数，λ 增大时，$\|\Gamma\|^2$ 递降，故可以用牛顿-拉夫生法找出 λ_{opt}。

根据需要，还可以有其他种约束条件与相应的有约束最小二乘估计。

第三章　随机变量估计

本章中将研究这样一类估计问题，即被估计量也是随机变量。

第一节介绍贝叶斯估计的一般原理，第二节和第三节则详细研究最大后验估计与最小方差估计，它们都需要关于概率密度函数的先验知识。接着研究两种更为实用的估计方法：线性最小方差估计与最小均方误差估计，最后，对第二、第三章中介绍的各种估计进行总结比较。

§3-1　贝叶斯估计原理

若观测量与被估计量都是随机量，并且已经知道它们的联合概率密度函数$p(z, x)$，则可以用贝叶斯估计原理对被估计量作出估计。

设被估计量为 x，其估计值为$\hat{x}$，则估计误差为

$$\varepsilon = \hat{x} - x \tag{3-1-1}$$

由于估计误差的存在而要付出代价，我们用代价函数$C(\varepsilon)$来表达代价与误差大小的关系，常用的代价函数有下列三种形式，图(3-1-1)中示出了三种代价函数的形状。

（一）均匀代价函数

$$C(\varepsilon)=\begin{cases}0, & |\varepsilon| \leqslant \dfrac{\delta}{2} \\ 1, & |\varepsilon| > \dfrac{\delta}{2}\end{cases} \quad \delta \to 0 \tag{3-1-2}$$

意指除了$\varepsilon=0$不要付出代价之外，只要有误差，就得付出代价，且代价是均匀不变的。

（二）误差平方代价函数

$$C(\varepsilon)=\varepsilon^2 \tag{3-1-3}$$

意指代价与误差的平方成比例。

（三）误差绝对值代价函数

$$C(\varepsilon)=|\varepsilon| \tag{3-1-4}$$

意指代价与误差的绝对值成比例。

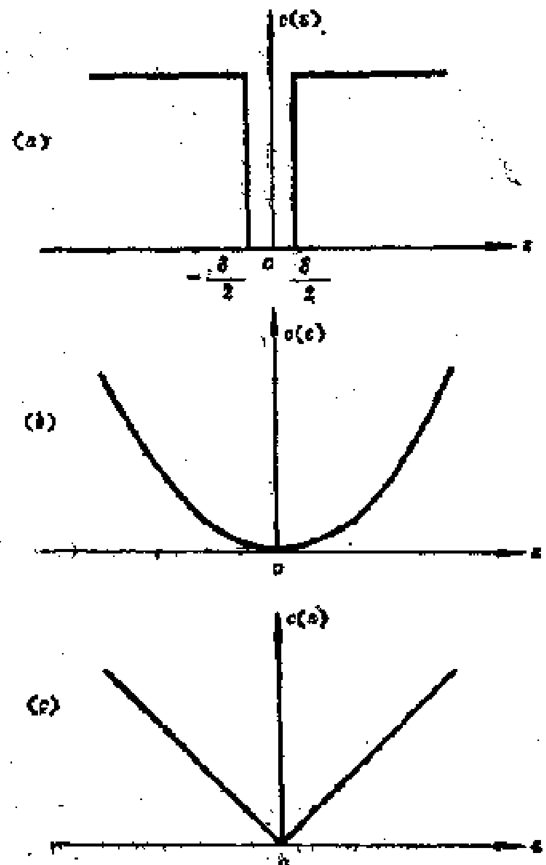


图3-1-1　三种代价函数形状

(a)均匀代价函数；(b)误差平方代价函数；(c)误差绝对值代价函数

在指定了代价函数与先验概率密度函数$p(z, x)$之后，就可以求出代价的均值，称为平均代价，其表达式为

$$\overline{C}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}C(\hat{x}-x)p(z,\ x)dzdx \qquad (3\text{-}1\text{-}5)$$

使平均代价$\overline{C}$达到最小时的$\hat{x}$值就称为 x的贝叶斯估计。

应用贝叶斯定理，$p(z,\ x)=p(x/z)p(z)$，把式(3-1-5)改写成

$$\overline{C}=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}C(\hat{x}-x)p(x/z)dx\right]p(z)dz \qquad (3\text{-}1\text{-}6)$$

在上式中，内积分和$p(z)$都是非负的，因此，使内积分

$$\overline{C}(\hat{x}/z)=\int_{-\infty}^{\infty}C(\hat{x}-x)p(x/z)dx \qquad (3\text{-}1\text{-}7)$$

达到最小即可使平均代价$\overline{C}$达到最小。$\overline{C}(\hat{x}/z)$称为条件平均代价，故根据使$\overline{C}(\hat{x}/z)$达到最小的条件即可求出最优估计$\hat{x}$。显然，当选用不同的代价函数时，得到的估计值也可能是不同的。

（一）对于均匀代价函数

记条件平均代价为$\overline{C}_{un}(\hat{x}/z)$，我们有

$$\overline{C}_{un}(\hat{x}/z)=1-\int_{\hat{x}_{un}-\frac{\delta}{2}}^{\hat{x}_{un}+\frac{\delta}{2}}p(x/z)dx \qquad (3\text{-}1\text{-}8)$$

为使$\overline{C}_{un}(\hat{x}/z)$达到极小，(3-1-8)式右边积分应为最大。令$\delta\to 0$，显见，以后验密度$p(x/z)$达到最大时的$x$值作为估计值，记作$\hat{x}_{map}$，称之为最大后验估计。

（二）对于误差平方代价函数

记条件平均代价为$\overline{C}_{mv}(\hat{x}/z)$，有

$$\overline{C}_{mv}(\hat{x}/z)=\int_{-\infty}^{\infty}(\hat{x}-x)^2p(x/z)dx \qquad (3\text{-}1\text{-}9)$$

根据定义，显见$\overline{C}_{mv}(\hat{x}/z)$乃是估计误差方差表达式，即$\overline{C}_{mv}(\hat{x}/z)=E[(\hat{x}-x)^2]$。令(3-1-9)式对$\hat{x}$求导数，并使之等于零，得

$$\frac{d}{d\hat{x}}\int_{-\infty}^{\infty}(\hat{x}-x)^2p(x/z)dx$$

$$=-2\int_{-\infty}^{\infty}xp(x/z)dx+2\hat{x}\int_{-\infty}^{\infty}p(x/z)dx=0 \qquad (3\text{-}1\text{-}10)$$

由此解出

$$\hat{x}_{mv}=\frac{\int_{-\infty}^{\infty}xp(x/z)dx}{\int_{-\infty}^{\infty}p(x/z)dx}=\int_{-\infty}^{\infty}xp(x/z)dx \qquad (3\text{-}1\text{-}11)$$

上式表明，估计$\hat{x}_{mv}$乃是后验密度$p(x/z)$的均值，故称$\hat{x}_{mv}$为后验密度均值估计。又根据(3-1-9)式的含义，$\overline{C}_{mv}(\hat{x}/z)$达到最小，也就是方差达到最小，故又称$\hat{x}_{mv}$为最小方差估计，即$\hat{x}_{mv}$是在一切估计中方差最小的估计。

（三）对于绝对值代价函数

记条件代价为$\overline{C}_{abs}(\hat{x}/z)$，有

$$\overline{C}_{abs}(\hat{x}/z)=\int_{-\infty}^{\infty}|\hat{x}-x|\,p(x/z)dx$$
$$=\int_{-\infty}^{\hat{x}}(x-\hat{x})p(x/z)dx$$
$$+\int_{\hat{x}}^{\infty}(\hat{x}-x)p(x/z)dx \tag{3-1-12}$$

两边对$\hat{x}$求导数，并使之等于零，得

$$\int_{-\infty}^{\hat{x}_{abs}}p(x/z)dx-\int_{\hat{x}_{abs}}^{\infty}p(x/z)dx=0 \tag{3-1-13}$$

可见，$\hat{x}_{abs}$乃是后验密度$p(x/z)$的中位数，所以$\hat{x}_{abs}$称为后验密度$p(x/z)$的中位数估计。

从以上三种估计的定义可以推知，如果后验密度$p(x/z)$具有单峰和对称形状，显然有

$$\hat{x}_{map}=\hat{x}_{ms}=\hat{x}_{abs}$$

当$p(x/z)$不具有上述性质时，三种估计可能是各不相同的，如图3-1-2所示。

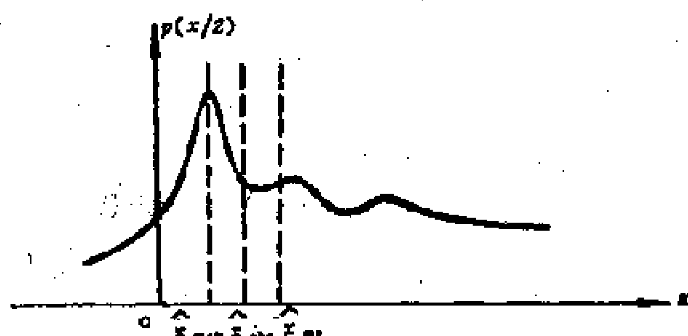


图3-1-2　三种贝叶斯估计

最小方差估计$\hat{x}_{ms}$与最大后验估计$\hat{x}_{map}$十分有用，以后常用到，而$\hat{x}_{abs}$较少用到。下面对$\hat{x}_{map}$和$\hat{x}_{ms}$的具体求法作进一步讨论。

§3-2　最大后验估计$\hat{X}_{map}$

根据上述关于最大后验估计的定义得知，$\hat{x}_{map}$对应于后验密度$p(x/z)$的极大值，于是得出表达式

$$\frac{\partial}{\partial x}p(x/z)\Big|_{x=\hat{x}_{map}}=0 \tag{3-2-1}$$

上式称为最大后验估计方程，其等价形式为

$$\frac{\partial}{\partial x}\ln p(x/z)\Big|_{x=\hat{x}_{map}}=0 \tag{3-2-2}$$

应当注意，当$p(x/z)$具有多个峰值时，应选用其中最大者作为$\hat{x}_{map}$求的条件，根据贝叶斯原理，把$p(x/z)$写成下式

$$p(x/z)=\frac{p(z/x)p(x)}{p(z)} \tag{3-2-3}$$

两边取对数，得

$$\ln p(x/z)=\ln p(z/x)+\ln p(x)-\ln p(z) \tag{3-2-4}$$

代入(3-2-2)式，得

$$\frac{\partial}{\partial x}\ln p(x/z)\Big|_{x=\hat{x}_{map}}$$

$$=\frac{\partial}{\partial x}\ln p(z/x)\Big|_{x=\hat{x}_{map}}+\frac{\partial}{\partial x}\ln p(x)\Big|_{x=\hat{x}_{map}}=0 \tag{3-2-5}$$

可以把最大似然估计看作是最大后验估计的一个特例。考察（3-2-5）式，若x为非随机量时，则$\frac{\partial \ln p(x)}{\partial x}=0$，于是有

$$\frac{\partial}{\partial x}\ln p(x/z)\Big|_{x=\hat{x}_{map}}=\frac{\partial}{\partial x}\ln p(z/x)\Big|_{x=\hat{x}_{ml}}=0$$

也就是说，最大似然估计是当x为非随机变量这一特殊情况下的最大后验估计。

下面研究几个求$\hat{x}_{map}$的例子：

例3-2-1　设有测量序列$z_i=x_i+e_i$，$i=1,2,\cdots,N$，其中x为高斯变量$N(0,\sigma_x^2)$，噪声e_i互相独立，均为高斯变量$N(0,\sigma_e^2)$，求最大后验估计$\hat{x}_{map}$。

解：记$Z^z=\{Z_i\}$，$i=1,2,\cdots,N$，根据题意，可写出条件密度函数$p(Z^z/x)$。

$$p(Z^z/x)=\prod_{i=1}^{N}\frac{1}{\sigma_e\sqrt{2\pi}}\exp\left[-\frac{(Z_i-x)^2}{2\sigma_e^2}\right] \tag{3-2-6}$$

和

$$p(x)=\frac{1}{\sigma_x\sqrt{2\pi}}\exp\left[-\frac{x^2}{2\sigma_x^2}\right] \tag{3-2-7}$$

代入贝叶斯公式，得

$$p(x/Z^z)=\frac{1}{p(z)}\left[\left(\prod_{i=1}^{N}\frac{1}{\sigma_e\sqrt{2\pi}}\right)\left(\frac{1}{\sigma_x\sqrt{2\pi}}\right)\right]$$

$$\times\exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_e^2}\sum_{i=1}^{N}(Z_i-x)^2+\frac{1}{\sigma_x^2}x^2\right]\right\} \tag{3-2-8}$$

对指数项进行配方，并把仅与Z_i有关的项归并成常数项$K(Z)$，得

$$p(x/Z^z)=K(Z^z)\exp\left\{-\frac{1}{2\sigma_z^2}\left[x-\frac{\sigma_x^2}{\sigma_x^2+\frac{\sigma_e^2}{N}}\left(\frac{1}{N}\sum_{i=1}^{N}Z_i\right)\right]^2\right\} \tag{3-2-9}$$

其中$K(Z^z)$代表不包含有x的系数项。

其中

$$\sigma_z^2=\left(\frac{1}{\sigma_x^2}+\frac{N}{\sigma_e^2}\right)^{-1}=\frac{\sigma_x^2\sigma_e^2}{N\sigma_x^2+\sigma_e^2} \tag{3-2-10}$$

记　$\rho=\frac{\sigma_e^2}{\sigma_x^2}$，$\rho$代表功率信噪比倒数。

由(3-2-9)式可知，后验密度$p(x/Z^z)$为高斯分布，其最大值必位于均值$E[x/Z^z]$处，于是可以得出

$$\hat{x}_{map}=E[x/Z^z]$$

$$=\frac{\sigma_x^2}{\sigma_x^2+\frac{\sigma_e^2}{N}}\left(\frac{1}{N}\sum_{i=1}^{N}Z_i\right)$$

$$=\frac{1}{N+\rho}\sum_{i=1}^{N}Z_i \tag{3-2-11}$$

其方差为

$$P_{map}=\sigma_{\hat{x}}^2=\frac{\sigma_e^2}{N+\rho} \tag{3-2-12}$$

考察(3-2-11)和(3-2-12)式，若信噪比很大，$\rho\approx 0$，则有 $\hat{x}_{map}\approx\frac{1}{N}\sum_{i=1}^{N}Z_i$，$P_{map}\approx\frac{1}{N}\sigma_e^2$，这时最大后验估计近似于平均值估计，这时就只需用测量值来求估计，而不必考虑$p(x)$。反之，若σ_e^2很大，而σ_x^2很小，则只需要相信先验知识$p(x)$，无需利用测量值。

例3-2-2　设有单次观测值$Z=x+e$，其中误差e为高斯分布$N(0,\sigma_e^2)$，而x为瑞利分布，方差为σ_x^2，求最大后验估计$\hat{x}_{map}$。

解：由噪声e的高斯性质，得

$$p(Z/x)=\frac{1}{\sigma_e\sqrt{2\pi}}\exp\left[-\frac{(Z-x)^2}{2\sigma_e^2}\right] \tag{3-2-13}$$

而

$$p(x)=\begin{cases}0 & x<0\\ \frac{x}{\sigma_x^2}e^{-\frac{x^2}{2\sigma^2 x}} & x\geqslant 0\end{cases} \tag{3-2-14}$$

代入最大后验估计方程中，得

$$\left\{\frac{\partial}{\partial x}\left[-\frac{(Z-x)^2}{2\sigma_e^2}\right]+\frac{\partial}{\partial x}\left[\ln x\quad\frac{x^2}{2\sigma_x^2}\right]\right\}_{x=\hat{x}_{map}}=0 \tag{3-2-15}$$

或

$$\left(1+\frac{\sigma_e^2}{\sigma_x^2}\right)\hat{x}_{map}\quad Z\hat{x}_{map}-\sigma_e^2=0$$

令

$$a=1+\frac{\sigma_e^2}{\sigma_x^2}$$

解出最大后验估计$\hat{x}_{map}$为

$$\hat{x}_{map}=\frac{Z}{2a}\left(1+\sqrt{\frac{4a\sigma_e^2}{Z^2}+1}\right) \tag{3-2-16}$$

这个结果表明，$\hat{x}_{map}$是测量值Z的非线性函数，仅当$Z\gg 2\sigma_e\sqrt{a}$时，$\hat{x}_{map}$才与Z成近似线性关系。

例3-2-3　设有测量值

$$z_i=x^3+e_i\qquad i=1,2,\cdots,N$$

其中x和e_i均为高斯分布，互相独立，噪声的分布为$N(0,\sigma_e^2)$；x的分布为$N(0,\sigma_x^2)$，求最大后验估计$\hat{x}_{map}$。

解：记$Z^k=\{z_i\}$，$i=1,2,\cdots,N$，由已知条件得

$$p(Z^k/x)=\frac{1}{(2\pi)^{\frac{N}{2}}\sigma_e^N}\cdot\exp\left[-\frac{1}{2\sigma_e^2}\sum_{i=1}^{N}(z_i-x^3)^2\right] \tag{3-2-17}$$

$$p(x)=\frac{1}{\sigma_x\sqrt{2\pi}}\exp\left[-\frac{x^2}{2\sigma_x^2}\right] \tag{3-2-18}$$

代入最大后验估计方程，得

$$\frac{\partial}{\partial x}\left[\sum_{i=1}^{N}\frac{(z_i-x^3)^2}{2\sigma_v^2}+\frac{x^2}{2\sigma_x^2}\right]\Bigg|_{x=\hat{x}_{map}}=0$$

或
$$\left[\sum_{i=1}^{N}\frac{-3x^2(z_i-x^3)}{\sigma_v^2}+\frac{x}{\sigma_x^2}\right]\Bigg|_{x=\hat{x}_{map}}=0$$

或
$$\frac{3N}{\sigma_v^2}\hat{x}_{map}-\left(3\sum_{i=1}^{N}\frac{z_i}{\sigma_v^2}\right)\hat{x}^2_{map}+\frac{1}{\sigma_x^2}\hat{x}_{map}=0 \qquad (3\text{-}2\text{-}19)$$

解此四次方程可得出$\hat{x}_{map}$，注意到该方程可能有四个解，应当检验一下，只有与$p(x/z)$的最大值对应的解才是需要的$\hat{x}_{map}$。

从以上两个例子可以看到，对于非高斯分布，求$\hat{x}_{map}$比较困难，而求计算误差方差将更加困难，类似于上章中求$\hat{x}_{ml}$（最大似然估计）的方差下界一样，也可以求$\hat{x}_{map}$的方差下界，现在来证明下述定理：

定理 若x是随机变量，对于任一无偏估计$\hat{x}_{map}$存在下列不等式

$$E\left[(\hat{x}_{map}-x)^2\right]\geqslant\frac{1}{E\left[\left(\frac{\partial\ln p(z,x)}{\partial x}\right)^2\right]} \qquad (3\text{-}2\text{-}20)$$

或其等价形式

$$E\left[(\hat{x}_{map}-x)^2\right]\geqslant\frac{1}{-E\left[\frac{\partial^2\ln p(z,x)}{\partial x^2}\right]} \qquad (3\text{-}2\text{-}21)$$

注意到这里$p(z,x)$是联合密度，取期望应在两个域上求平均。而且是，当且仅当

$$\frac{\partial}{\partial x}\ln p(z,x)=k(\hat{x}_{map}-x) \qquad k=\text{常数} \qquad (3\text{-}2\text{-}22)$$

对一切z和x均成立时，不等式才能取等号。若（3-2-22）式得到满足，则此无偏估计的方差将达到下界，此下界称为**随机变量的克拉美-罗方差下界**，此无偏估计称为**优效估计**。

对于多次测量，令$Z^k=\{Z_i\}, i=1,2,\cdots,N$，对应的克拉美-罗不等式为

$$E\left[(\hat{x}_{map}-x)^2\right]\geqslant\frac{1}{E\left[\left(\frac{\partial}{\partial x}\ln p(Z^k,x)\right)^2\right]} \qquad (3\text{-}2\text{-}23)$$

或
$$E\left[(\hat{x}_{map}-x)^2\right]\geqslant\frac{1}{-E\left[\frac{\partial^2}{\partial x^2}\ln p(Z^k,x)\right]} \qquad (3\text{-}2\text{-}24)$$

现在来证明上述定理，令误差的条件均值为$B(x)$

$$B(x)=\int_{-\infty}^{\infty}(\hat{x}-x)p(z/x)dz \qquad (3\text{-}2\text{-}25)$$

将上式两边乘以$p(x)$，然后对x求导数，得

$$\frac{d}{dx}p(x)B(x)=-\int_{-\infty}^{\infty}p(z,x)dz+\int_{-\infty}^{\infty}\frac{\partial p(z,x)}{\partial x}(\hat{x}-x)dz \qquad (3\text{-}2\text{-}26)$$

再对x求积分，得

$$p(x)B(x)\Big|_{-\infty}^{\infty} = -1+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\partial p(z,x)}{\partial x}(\hat{x}-x)dxdz \tag{3-2-27}$$

因为
$$\lim_{x\to\infty} B(x)p(x)=0$$
$$\lim_{x\to-\infty} B(x)p(x)=0$$

故(3-2-27)式成为

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\partial p(z,x)}{\partial x}(\hat{x}-x)dxdz=1 \tag{3-2-28}$$

采用求$\hat{x}_{ml}$的方差下界同样的方法，应用施瓦兹不等式，可以得到(3-2-20)和(3-2-21)式

不等式取等号的条件是

$$\frac{\partial}{\partial x}\ln p(z,x)=k(\hat{x}_{map}-x) \qquad k=\text{常数} \tag{3-2-29}$$

对上式再次求导，得出等价条件为

$$\frac{\partial^2}{\partial x^2}\ln p(z,x)=-k \tag{3-2-30}$$

由此可见，不等式取等号的条件是要求$\ln p(z,x)$是x的二次函数，或者说，要求

$$\ln p(z,x)=ax^2+bx+c \tag{3-2-31}$$

据此，可以得出结论：只有当$p(z,x)$为高斯分布时，(3-2-31)式才能满足。最大后验估计$\hat{x}_{map}$才能成为优效估计，其方差才能达到克拉美-罗下界。

当X为矢量情况时，推导矢量$\hat{X}_{map}$方差下界的方法与推导$\hat{X}_{ml}$方差下界相似。这时，费什尔信息矩阵由两部分组成。

$$J_{map}=J_{ml}+J_x \tag{3-2-32}$$

其中
$$J_{ml}=E\left[\left(\frac{\partial \ln p(Z/X)}{\partial X}\right)\left(\frac{\partial \ln p(Z/X)}{\partial X}\right)^T\right]$$
$$=-E\left[\frac{\partial}{\partial X}\left(\frac{\partial}{\partial X}\ln p(Z/X)\right)^T\right] \tag{3-2-33}$$
$$J_x=E\left[\left(\frac{\partial p(X)}{\partial X}\right)\left(\frac{\partial p(X)}{\partial X}\right)^T\right]$$
$$=-E\left[\frac{\partial}{\partial X}\left(\frac{\partial}{\partial X}p(X)\right)^T\right] \tag{3-2-34}$$

信息矩阵J_{map}^{-1}的对角线元素是对应的$\hat{x}_i$的方差下界，当$p(X/Z)$对一切Z均为高斯形时，则J_{map}为常数矩阵，此时$E\left[(\hat{X}_{map}-X)(\hat{X}_{map}-X)^T\right]=J_{map}^{-1}$，所有估计量均是优效估计量。

§3-3 最小方差估计$\hat{X}_{mv}$

在本章第一节中，已证明了最小方差估计$\hat{x}_{mv}$乃是后验密度$p(x/z)$的均值。

$$\hat{x}_{mv}=E[x/z]=\frac{\int_{-\infty}^{\infty}xp(x/z)dx}{\int_{-\infty}^{\infty}p(x/z)dx}=\int_{-\infty}^{\infty}xp(x/z)dx \tag{3-3-1}$$

现在来研究一下$\hat{x}_{mv}$的性质。

1. $\hat{x}_{mv}$的均值$E[\hat{x}_{mv}]$

$$\begin{aligned}E[\hat{x}_{mv}]&=\int_{-\infty}^{\infty}E[x/z]p(z)dz\\&=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}xp(x/z)dx\right]p(z)dz\\&=\int_{-\infty}^{\infty}x\left[\int_{-\infty}^{\infty}p(z,x)dz\right]dx\\&=\int_{-\infty}^{\infty}xp(x)dx=E[x]\end{aligned} \tag{3-3-2}$$

可见，最小方差估计$\hat{x}_{mv}$是无偏估计，其均方误差就等于方差。

2. $\hat{x}_{mv}$的方差$E[(\hat{x}_{mv}-x)^2]$

$$\begin{aligned}\mathrm{var}[\hat{x}_{mv}-x]=E[(\hat{x}_{mv}-x)^2]&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\hat{x}_{mv}-x)^2p(z,x)dxdz\\&=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}[x-E(x/z)]^2p(x/z)dx\right]p(z)dz\\&=\int_{-\infty}^{\infty}\mathrm{var}[x/z]p(z)dz\end{aligned} \tag{3-3-3}$$

上式表明，$\mathrm{var}[\hat{x}_{mv}]$乃是$\mathrm{var}[x/z]$对$z$求平均。

根据$\hat{x}_{mv}$的定义，$\hat{x}_{mv}$是一切估计中方差最小的估计，由于$\hat{x}_{mv}$具有无偏，方差最小的性质，常称为最优估计。

3. 把$\hat{x}_{mv}$与$\hat{x}_{map}$作一比较。$\hat{x}_{mv}$是由后验密度$p(x/z)$的均值确定的，显然，对于高斯分布而言，由于其对称性和单峰性，必有

$$\hat{x}_{mv}=\hat{x}_{map} \tag{3-3-4}$$

下面来研究几个求$\hat{x}_{mv}$的例子。

例3-3-1 设有线性矢量测量方程

$$Z=HX+e \tag{3-3-5}$$

其中X为$N\times1$维矢量，Z和噪声e为$M\times1$维矢量，H为$M\times N$维测量矩阵，假定X与e独立，且均为高斯分布，求最小方差估计X_{mv}。已知

$$E[Z]=\mu \qquad \mathrm{var}[X]=P$$
$$E[e]=0 \qquad \mathrm{var}[e]=R$$

解：为了求得最小方差估计$\hat{X}_{mv}$，需要知道后验密度$p(X/Z)$，由贝叶斯公式得

$$p(X/Z)=\frac{p(Z/X)p(X)}{p(Z)}$$

式中所有密度函数都是高斯分布，其中

$$E[Z/X]=HX \qquad \mathrm{var}[Z/X]=R$$
$$E[Z]=H\mu \qquad \mathrm{var}[Z]=HPH^T+R$$

写出各密度函数如下：

$$p(Z/X)=\frac{1}{(2\pi)^{\frac{N}{2}}|R|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(Z-HX)^{T}R^{-1}(Z-HX)\right] \qquad (3\text{-}3\text{-}6)$$

$$p(X)=\frac{1}{(2\pi)^{\frac{N}{2}}|P|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(X-\mu)^{T}P^{-1}(X-\mu)\right] \qquad (3\text{-}3\text{-}7)$$

$$p(Z)=\frac{1}{(2\pi)^{\frac{N}{2}}|HPH^{T}+R|}$$

$$\times\exp\left[-\frac{1}{2}(Z-H\mu)^{T}(HPH^{T}+R)^{-1}(Z-H\mu)\right] \qquad (3\text{-}3\text{-}8)$$

于是，得到后验密度为

$$p(X/Z)=\frac{|HPH^{T}+R|^{\frac{1}{2}}}{(2\pi)^{\frac{N}{2}}|P|^{\frac{1}{2}}|R|^{\frac{1}{2}}}$$

$$\times\exp\left[-\frac{1}{2}(X-\hat{X}_{ms})^{T}P_{ms}^{-1}(X-\hat{X}_{ms})\right] \qquad (3\text{-}3\text{-}9)$$

其中

$$\hat{X}_{ms}=(P^{-1}+H^{T}R^{-1}H)^{-1}(H^{T}R^{-1}Z+P^{-1}\mu) \qquad (3\text{-}3\text{-}10)$$

$$P_{ms}=(P^{-1}+H^{T}R^{-1}H)^{-1} \qquad (3\text{-}3\text{-}11)$$

再对本例求最大后验估计，因为高斯密度函数的最大值必与均值对应，所以可立即得出结论

$$\hat{X}_{ms}=\hat{X}_{map}$$

(3-3-10)和(3-3-11)式可以改写成另一形式，通过矩阵運算，我們有

$$P_{ms}=(P^{-1}+H^{T}R^{-1}H)^{-1}$$

$$=P-PH^{T}(HPH^{T}+R)^{-1}HP \qquad (3\text{-}3\text{-}12)$$

记

$$K=PH^{T}(HPH^{T}+R)^{-1}$$

$$=(P^{-1}+H^{T}R^{-1}H)^{-1}H^{T}R^{-1}$$

$$=P_{ms}H^{T}R^{-1} \qquad (3\text{-}3\text{-}13)$$

代回(3-3-10)和(3-3-12)中，得

$$\hat{X}_{ms}=\mu+K(Z-H\mu) \qquad (3\text{-}3\text{-}14)$$

$$P_{ms}=(I-KH)P \qquad (3\text{-}3\text{-}15)$$

(3-3-13)、(3-3-14)、(3-3-15)三式构成求$\hat{X}_{ms}$和P_{ms}的一组算法，并与(3-3-10)、(3-3-11)式等价。

例3-3-2　一种比较常见的情况是待估计量x是标量，由多次测量值

$$z_i=x_i+e_i \qquad i=1,2,\cdots,N$$

求最小方差估计$\hat{x}_{ms}$。

解：写成矢量测量方程，得

$$Z=Hx+e \qquad (3\text{-}3\text{-}16)$$

其中

$$Z=\begin{pmatrix}z_1\\z_2\\\vdots\\z_N\end{pmatrix},\qquad H=\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix},\qquad e=\begin{pmatrix}e_1\\e_2\\\vdots\\e_N\end{pmatrix}$$

其中 x 为随机变量，假设

$$E[x]=\mu \qquad \operatorname{var}[x]=\sigma_x^2$$
$$E[e_i]=0 \qquad \operatorname{var}[e_i]=\sigma_e^2$$

代入(3-3-9)式，得

$$p_{ms}=\left(\frac{1}{\sigma_x^2}+\frac{N}{\sigma_e^2}\right)^{-1}$$

$$=\frac{\sigma_e^2\ \sigma_x^2}{N\sigma_x^2+\sigma_e^2}=\frac{\sigma_e^2}{N+\dfrac{\sigma_e^2}{\sigma_x^2}} \tag{3-3-17}$$

再由(3-3-10)式，得

$$\hat{x}_{ms}=P_{ms}\left(\frac{1}{\sigma_e^2}\sum_{i=1}^{N}z_i+\frac{\mu}{\sigma_x^2}\right)$$

$$=\frac{1}{1+N\dfrac{\sigma_x^2}{\sigma_e^2}}\left(\mu+\frac{\sigma_x^2}{\sigma_e^2}\sum_{i=1}^{N}z_i\right) \tag{3-3-18}$$

若 $\mu=0$，则(3-3-17)、(3-3-18)式与上一节中例3-2-1所得结果完全相同。

例3-3-3　再研究在非高斯分布情况下求最小方差估计的例子，设有单次测量值

$$z=x+e$$

其中 e 为零均值高斯噪声，而 x 在 $-s$ 与 $+s$ 之间均匀分布，求最小方差估计 $\hat{x}_{ms}$。

解：由已知条件，可得出

$$p(z/x)=\frac{1}{\sigma_e\sqrt{2\pi}}\exp\left[-\frac{(z-x)^2}{2\sigma_e^2}\right] \tag{3-3-19}$$

由贝叶斯公式

$$p(x/z)=\frac{p(z/x)p(x)}{p(z)}$$

得

$$\hat{x}_{ms}=\frac{\int_{-\infty}^{\infty}xp(x/z)dx}{\int_{-\infty}^{\infty}p(x/z)dx}$$

$$=\frac{\int_{-\infty}^{\infty}\dfrac{xp(x)p(z/x)}{p(z)}dx}{\int_{-\infty}^{\infty}\dfrac{p(x)p(z/x)}{p(z)}dx}$$

$$=\frac{\int_{-\infty}^{\infty}xp(x)p(z/x)dx}{\int_{-\infty}^{\infty}p(x)p(z/x)dx} \tag{3-3-20}$$

其中 $\int_{-\infty}^{\infty}p(x)p(z/x)dx=\dfrac{1}{2S}\int_{-s}^{s}\dfrac{1}{\sigma_e\sqrt{2\pi}}\left(e^{-\frac{(z-x)^2}{2\sigma_e^2}}\right)dx$　(3-3-21)

$$\int_{-\infty}^{\infty} x p(x) p(z/x) d x = \frac{1}{2S} \int_{-s}^{s} \frac{x}{\sigma_e \sqrt{2\pi}} \left(e^{-\frac{(z-x)^2}{2\sigma_e^2}} \right) d x \tag{3-3-22}$$

代入(3-3-20)式中，得

$$\hat{x}_{mv} = \frac{\int_{-s}^{s} x e^{-\frac{(z-x)^2}{2\sigma_e^2}} d x}{\int_{-s}^{s} e^{-\frac{(z-x)^2}{2\sigma_e^2}} d x} \tag{3-3-23}$$

(3-3-23)式十分复杂，下面推导近似值表达式，利用误差函数

$$\mathrm{erf}(y) = \frac{1}{\sqrt{2\pi}} \int_0^y e^{-\frac{x^2}{2}} d x$$

可以把$\hat{x}_{mv}$的表达式变成

$$\hat{x}_{mv} = z - \frac{\sigma_e}{\Phi(\mu)\sqrt{2\pi}} \left(e^{-\frac{(a-u)^2}{2}} - e^{-\frac{(a+u)^2}{2}} \right) \tag{3-3-24}$$

其中 $a = \dfrac{s}{\sigma_e}$ $\quad u = \dfrac{z}{\sigma_e}$

$$\Phi(u) = \mathrm{erf}(a-u) + \mathrm{erf}(a+u)$$

再记 $$g(u) = e^{-\frac{(a-u)^2}{2}} + e^{-\frac{(a+u)^2}{2}}$$

则有 $$\hat{x}_{mv} = z - \frac{\sigma_e}{\Phi(u)\sqrt{2\pi}} g(u) \tag{3-3-25}$$

由于$\hat{x}_{mv}$的表达式是高度非线性的，用近似式 简 化 之，当$a>3$时，$\mathrm{erf}(a) \approx 0.5$，因此$g(u)$可以近似为

$$g(u) = \begin{cases} 2 a u e^{-\frac{a^2}{2}} & |u| \ll a \text{或} |z| \ll s \\ 1 & |u| = a \\ e^{-\ [illegible]} & |u| \gg a \text{或} |z| \gg s \end{cases}$$

对于大的测量值$z \gg s$，则(3-3-23)式后一项可以略去，便得

$$\hat{x}_{mv} \approx s$$

图3-3-1中画出了函数$\Phi(u)$和$g(u)$的图形。

例3-3-4 设有二元通信系统，信号等概率地取值$\pm s$，有一组测量值

$$z_i = x_i + e_i \qquad i = 1, 2, \cdots, N$$

已知e_i为零均值高斯白噪声，方差为σ_e^2，求x的最小方差估计$\hat{x}_{mv}$。

解： 由已知条件写出x的密度函数

$$p(x) = \frac{1}{2} \delta(x+s) + \frac{1}{2} \delta(x-s)$$

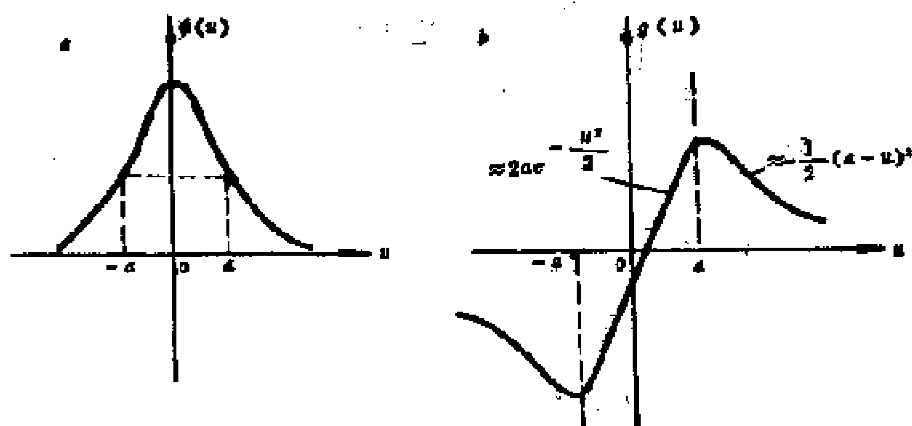


图3-3-1 $\Phi(u)$和$g(u)$图形

a—$\Phi(u)$； b—$g(u)$

其中$\delta(x)$为delta函数，记

$$e=\begin{pmatrix}e_1\\e_2\\\vdots\\e_N\end{pmatrix}\qquad Z=\begin{pmatrix}z_1\\z_2\\\vdots\\z_N\end{pmatrix}$$

写出密度函数$p(e)$和$p(z/x)$

$$p(e)=\left[\frac{1}{(2\pi\sigma_e^2)^{\frac{N}{2}}}\right]\exp\left[-\frac{1}{2\sigma_e^2}\sum_{i=1}^{N}e_i^2\right]\tag{3-3-26}$$

$$p(z/x)=\left[\frac{1}{(2\pi\sigma_e^2)^{\frac{N}{2}}}\right]\exp\left[-\frac{1}{2\sigma_e^2}\sum_{i=1}^{N}(z_i-x)^2\right]\tag{3-3-27}$$

因为

$$\hat{x}_{ms}=\frac{\int_{-\infty}^{\infty}xp(x/Z)dx}{\int_{-\infty}^{\infty}p(x/Z)dx}=\frac{\int_{-\infty}^{\infty}\frac{xp(Z/x)p(x)}{p(Z)}dx}{\int_{-\infty}^{\infty}\frac{p(Z/x)p(x)}{p(Z)}dx}$$

$$=\frac{\int_{-\infty}^{\infty}xp(x)p(Z/x)dx}{\int_{-\infty}^{\infty}p(x)p(Z/x)dx}\tag{3-3-28}$$

把相应的密度函数代入，得

$$\hat{x}_{ms}=\frac{p(Z/s)-p(Z/-s)}{p(Z/s)+p(Z/-s)}\tag{3-3-29}$$

又

$$\frac{p(Z/s)}{p(Z/-s)}=\exp\left[-\sum_{i=1}^{N}\frac{(z_i-s)^2}{2\sigma_e^2}+\sum_{i=1}^{N}\frac{(z_i+s)^2}{2\sigma_e^2}\right]$$

$$=\exp\left[\frac{2Ns\bar{z}}{\sigma_e^2}\right]\tag{3-3-30}$$

其中 $\bar{z}=\frac{1}{N}\sum_{i=1}^{N}z_i$

代入(3-3-29)式中，得

$$\hat{x}_{ms} = s\cdot \mathrm{th}\left(\frac{Ns}{\sigma_v^2}\bar{z}\right) = s\cdot \mathrm{th}(y) \tag{3-3-31}$$

其中 $y = \frac{Ns}{\sigma_v^2}\bar{z}$

$\mathrm{th}(y)$ 表示双曲正切函数，它具有下列性质

$$\mathrm{th}(y) \approx \begin{cases} y, & |y| \ll 1 \\ 1, & y \gg 1 \\ -1, & y \ll -1 \end{cases}$$

$\mathrm{th}(y)$ 函数的图形如图3-3-2所示。

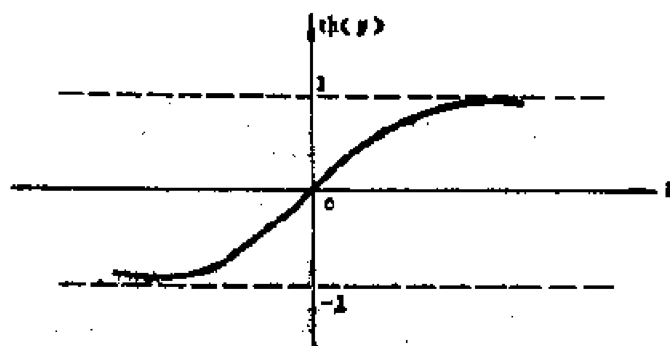


图3-3-2　双曲正切函数图形

例3-3-5　某一实验的事件数 n 服从泊松分布，泊松分布的均值为 a，于是有

$$p_r(n/a) = \frac{a^n}{n!}e^{-a} \qquad n = 0, 1, 2, \cdots \tag{3-3-32}$$

假定 a 的密度函数为

$$p(a) = \begin{cases} \lambda e^{-\lambda a} & a > 0 \\ 0 & a \leqslant 0 \end{cases} \tag{3-3-33}$$

要求由事件数 n 来估计参数 a。

解： 因为 a 的后验密度为

$$p(a/n) = \frac{p_r(n/a)p(a)}{p_r(n)} \tag{3-3-34}$$

把(3-3-32)、(3-3-33)式代入(3-3-34)式中，得

$$p(a/n) = k(n)a^n e^{-a(1+\lambda)} \tag{3-3-35}$$

其中　$$k(n) = \frac{(1+\lambda)^{n+1}}{n!} \tag{3-3-36}$$

于是参数 a 的最小方差估计为

$$\hat{a}_{ms} = \int_0^\infty a\,p(a/n)\,da$$

即有

$$\hat{a}_{ms} = \frac{(1+\lambda)^{n+1}}{n!}\int_0^\infty a^{n+1}e^{-a(1+\lambda)}\,da$$

$$= \frac{(1+\lambda)^{n+1}}{(1+\lambda)^{n+2}}(n+1) = \frac{1}{1+\lambda}(1+n) \tag{3-3-37}$$

再求a的最大后验估计$\hat{a}_{map}$,对(3-3-35)式取对数，并代入最大后验估计方程，得

$$\ln p(a/n)=n\ln a-a(1+\lambda)+\ln k(n)$$

$$\frac{\partial}{\partial a}\ln p(a/n)\bigg|_{a=\hat{a}_{map}}=0 \tag{3-3-38}$$

可以解出

$$\hat{a}_{map}=\frac{n}{1+\lambda} \tag{3-3-39}$$

显见，$\hat{a}_{map}$不等于$\hat{a}_{mv}$。一般地说，后验密度为非高斯分布时，最大后验估计不等于最小方差估计。

§3-4 线性最小方差估计$\hat{X}_{lmv}$

最小方差估计$\hat{X}_{mv}$具有无偏、方差最小的优点，是一切估计中最好的估计，但是从上面讨论中我们看到，在许多实际情况下$\hat{X}_{mv}$是很难求得的，需要的先验知识太多，实际上又往往不具备，所得到的估计$\hat{X}_{mv}$与测量值Z常常成强非线性关系,很少得到简便的结果，应用起来很不方便，因此，应寻找次优估计方法，线性最小方差估计$\hat{X}_{lmv}$就是其中的一种。

假定并不知道联合密度函数$p(Z,X)$的完整知识，但知道关于X和Z的头两阶矩，即已知$E[X]$和$E[Z]$，var$[X]$和var$[Z]$以及协方差cov$[X,Z]$。

现在规定估计$\hat{X}_{lmv}$是观测集的线性函数，即令

$$\hat{X}_{lmv}=A+BZ \tag{3-4-1}$$

若X为标量，Z为M维矢量，则A为标量，B为M维行矢量；若X为N维矢量，则A也为N维矢量，B为$N\times M$维矩阵。A和B均为待定量。

确定系数A和B采用下述准则：即使估计误差平方和的均值达到最小，这个准则通常称为均方误差最小准则。

若x为标量，均方误差为

$$J=E[(x-\hat{x})^2]=E[(x-A-BZ)^2] \tag{3-4-2}$$

若X为矢量，均方误差为

$$\begin{aligned}J&=E[(X-\hat{X})^T(X-\hat{X})]=\mathrm{tr}E[(X-\hat{X})(X-\hat{X})^T]\\&=E[(X-A-BZ)^T(X-A-BZ)]\\&=\mathrm{tr}E[(X-A-BZ)(X-A-BZ)^T]\end{aligned} \tag{3-4-3}$$

矢量情况更有一般性，我们来研究如何求线性最小方差估计$\hat{X}_{lmv}$。令J分别对A和B求偏导数，并使之等于零

$$\frac{\partial}{\partial A}E[(X-A-BZ)^T(X-A-BZ)]$$

$$= E\frac{\partial}{\partial \boldsymbol{A}}[(\boldsymbol{X}-\boldsymbol{A}-\boldsymbol{BZ})^T(\boldsymbol{X}-\boldsymbol{A}-\boldsymbol{BZ})]$$
$$= -2E[(\boldsymbol{X}-\boldsymbol{A}-\boldsymbol{BZ})]$$
$$= 2(\boldsymbol{A}+\boldsymbol{B}E\boldsymbol{Z}-E\boldsymbol{X})$$
$$= 2(\boldsymbol{A}+\boldsymbol{B}\overline{\boldsymbol{Z}}-\overline{\boldsymbol{X}})=0 \tag{3-4-4}$$

解出
$$\boldsymbol{A}=\overline{\boldsymbol{X}}-\boldsymbol{B}\overline{\boldsymbol{Z}} \tag{3-4-5}$$

$$\frac{\partial}{\partial \boldsymbol{B}}[(\boldsymbol{X}-\boldsymbol{A}-\boldsymbol{BZ})^T(\boldsymbol{X}-\boldsymbol{A}-\boldsymbol{BZ})]$$
$$= E[\frac{\partial}{\partial \boldsymbol{B}}(\boldsymbol{X}-\boldsymbol{A}-\boldsymbol{BZ})^T(\boldsymbol{X}-\boldsymbol{A}-\boldsymbol{BZ})]$$
$$= -2E[(\boldsymbol{X}-\boldsymbol{A}-\boldsymbol{BZ})\boldsymbol{Z}^T]$$
$$= 2[(\boldsymbol{A}\overline{\boldsymbol{Z}}^T+\boldsymbol{B}E[\boldsymbol{ZZ}^T]-E(\boldsymbol{XZ}^T)]=0 \tag{3-4-6}$$

把(3-4-5)代入(3-4-6)式，得
$$\boldsymbol{B}[E[\boldsymbol{ZZ}^T]-[\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}^T]-E[\boldsymbol{XZ}^T]+\overline{\boldsymbol{X}}\overline{\boldsymbol{Z}}^T=0 \tag{3-4-7}$$

记
$$\mathrm{var}(\boldsymbol{Z})=E[\boldsymbol{ZZ}^T]-\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}^T \tag{3-4-8}$$
$$\mathrm{cov}(\boldsymbol{XZ})=E[\boldsymbol{XZ}^T]-\overline{\boldsymbol{X}}\,\overline{\boldsymbol{Z}}^T \tag{3-4-9}$$

代入(3-4-7)式，得
$$\boldsymbol{B}\,\mathrm{var}(\boldsymbol{Z})-\mathrm{cov}(\boldsymbol{X},\boldsymbol{Z})=0$$
$$\boldsymbol{B}=\mathrm{cov}(\boldsymbol{X},\boldsymbol{Z})[\mathrm{var}\boldsymbol{Z}]^{-1} \tag{3-4-10}$$

于是
$$\hat{\boldsymbol{X}}_{lmv}=\boldsymbol{A}+\boldsymbol{BZ}=\overline{\boldsymbol{X}}-\boldsymbol{B}\overline{\boldsymbol{Z}}+\boldsymbol{BZ}$$
$$=\overline{\boldsymbol{X}}+\mathrm{cov}(\boldsymbol{X},\boldsymbol{Z})(\mathrm{var}\boldsymbol{Z})^{-1}(\boldsymbol{Z}-\overline{\boldsymbol{Z}}) \tag{3-4-11}$$

只要知道关于 $\boldsymbol{X}$ 和 $\boldsymbol{Z}$ 的头两阶矩，就可由(3-4-11)式求出 $\hat{\boldsymbol{X}}_{lmv}$。再来求估计误差方差阵 $\boldsymbol{P}_{lmv}$，估计误差为

$$\tilde{\boldsymbol{X}}=\boldsymbol{X}-\hat{\boldsymbol{X}}_{lmv}$$
$$=(\boldsymbol{X}-\overline{\boldsymbol{X}})-\mathrm{cov}(\boldsymbol{X},\boldsymbol{Z})(\mathrm{var}\boldsymbol{Z})^{-1}(\boldsymbol{Z}-\overline{\boldsymbol{Z}})$$

估计误差方差阵为

$$\boldsymbol{P}_{lmv}=E(\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^T)$$
$$=E[\{(\boldsymbol{X}-\overline{\boldsymbol{X}})-\mathrm{cov}(\boldsymbol{X},\boldsymbol{Z})(\mathrm{var}\boldsymbol{Z})^{-1}(\boldsymbol{Z}-\overline{\boldsymbol{Z}})\}$$
$$\times\{(\boldsymbol{X}-\overline{\boldsymbol{X}})-\mathrm{cov}(\boldsymbol{X},\boldsymbol{Z})(\mathrm{var}\boldsymbol{Z})^{-1}(\boldsymbol{Z}-\overline{\boldsymbol{Z}})\}^T] \tag{3-4-12}$$

其中 $\mathrm{var}\boldsymbol{Z}$ 为正定对称方阵，其转置等于本身，最后得估计误差方差阵为

$$\boldsymbol{P}_{lmv}=\mathrm{var}\,\boldsymbol{X}-\mathrm{cov}(\boldsymbol{X},\boldsymbol{Z})(\mathrm{var}\boldsymbol{Z})^{-1}\mathrm{cov}(\boldsymbol{Z},\boldsymbol{X}) \tag{3-4-13}$$

下面我们来研究线性最小方差估计的性质。

1. 无偏性。对 $\hat{\boldsymbol{X}}_{lmv}$ 取期望，得

$$E[\hat{\boldsymbol{X}}_{lmv}]=E\boldsymbol{X}+\mathrm{cov}(\boldsymbol{X},\boldsymbol{Z})(\mathrm{var}\boldsymbol{Z})^{-1}(E\boldsymbol{Z}-E\boldsymbol{Z})=E\boldsymbol{X} \tag{3-4-14}$$

所以，$\hat{\boldsymbol{X}}_{lmv}$ 是无偏估计。

2. 方差阵 $\boldsymbol{P}_{lmv}$。由于 $\hat{\boldsymbol{X}}$ 是无偏估计，均方误差最小也就是方差最小，所以 $\hat{\boldsymbol{X}}_{lmv}$ 是一切线性估计中方差最小的估计，因此，$\hat{\boldsymbol{X}}_{lmv}$ 常称为线性无偏最小方差估计。

3. 对于高斯分布情况，头两阶矩表达了概率密度的全部信息，可以推知，线性最小方差估计将等于最小方差估计，反过来还可以推知，如果最小方差估计表达成测量集的线性函数，那末它必定又是线性最小方差估计。

下面研究两个求 $\hat{X}_{lmv}$ 的例子。

例3-4-1　设有矢量线性测量方程

$$Z = HX + e$$

已知　$EX = \mu$，$\mathrm{var}\,X = E[(X - \bar{X})(X - \bar{X})^T] = P$

$Ee = 0 \quad \mathrm{var}\,e = R \quad E[Xe^T] = 0$

求线性最小方差估计 $\hat{X}_{lmv}$。

解：根据已知条件，得

$$EZ = HEX = H\mu \tag{3-4-15}$$

$$\mathrm{var}Z = E[[H(X-\mu)+e][H(X-\mu)+e]^T]$$
$$= HPH^T + R \tag{3-4-16}$$

$$\mathrm{cov}(X, Z) = \mathrm{cov}(Z, X)^T$$
$$= E[(X-\mu)(H(X-\mu)+e)^T]$$
$$= PH^T \tag{3-4-17}$$

代入(3-4-11)式，得

$$\hat{X}_{lmv} = \mu + PH^T(HPH^T + R)^{-1}(Z - H\mu) \tag{3-4-18}$$

代入(3-4-12)式得

$$P_{lmv} = P - PH^T(HPH^T + R)^{-1}HP \tag{3-4-19}$$

记　$$K = PH^T(HPH^T + R)^{-1} \tag{3-4-20}$$

则有　$$\hat{X}_{lmv} = \mu + K(z - H\mu)$$

$$P_{lmv} = (I - KH)P \tag{3-4-22}$$

把本例与例3-3-1比较，可以得出结论如下：在线性测量方程情况下，且联合密度 $P(X, Z)$ 为高斯分布，则最小方差估计与线性最小方差估计一致，即有

$$\hat{X}_{mv} = \hat{X}_{lmv}, \qquad P_{mv} = P_{lmv}$$

但是，在非高斯情况下，求 $\hat{X}_{lmv}$ 比求 $\hat{X}_{mv}$ 要简便得多，为了便于比较起见，仍以例3-3-3为例，求线性最小方差估计。

例3-4-2　设有单次测量 $z = x + e$，e 为零均值高斯噪声 $N(0, \sigma_e^2)$，x 在 $-s$ 与 $+s$ 之间均匀分布，求 $\hat{X}_{lmv}$ 及 P_{lmv}。

解：由题意可知

$$E[x] = 0 \qquad E[z] = 0$$

$$\mathrm{var}x = E[(x - \bar{x})^2] = \frac{1}{2s}\int_{-s}^{+s} x^2 dx = \frac{s^2}{3}$$

$$\mathrm{var}z = E[(z - \bar{z})^2] = \frac{s^2}{3} + \sigma_e^2$$

$$\mathrm{cov}(x, z) = E[(x - \bar{x})(z - \bar{z})] = \frac{s^2}{3}$$

于是，得到

$$\hat{x}_{lmv} = \frac{\frac{s^2}{3} z}{\frac{s^2}{3} + \sigma_v^2} = \frac{z}{1 + \frac{3\sigma_v^2}{s^2}}$$

$$P_{lmv} = \frac{s^2}{3} - \frac{\left(\frac{s^2}{3}\right)^2}{\frac{s^2}{3} + \sigma_v^2} = \frac{\frac{s^2}{3} \cdot \sigma_v^2}{\frac{s^2}{3} + \sigma_v^2} = \frac{\sigma_v^2}{1 + \frac{3\sigma_v^2}{s^2}}$$

其中$\frac{s^2}{\sigma_v^2}$为功率信噪比，在大功率信噪比时，有$\hat{X}_{lmv} \approx z$，$P_{lmv} \approx \sigma_v^2$。

例3-4-3 研究一个离散概率分布的例子，其分布情况如下：

若测量值$z = -1$，有$P(-1) = \frac{1}{10}$，$P(0) = \frac{2}{10}$，其它为零；若测量值$z = 0$，有$p(0) = \frac{1}{10}$，$P(1) = \frac{3}{10}$，其它为零；若量测值 $z = 1$，有 $P(1) = \frac{1}{10}$，$P(2) = \frac{2}{10}$，其它为零。求线性最小方差估计$\hat{x}_{lmv}$和最小方差估计$\hat{x}_{mv}$。

解：先求$\hat{x}_{lmv}$。根据已知条件，得

$$E[x] = \sum_{1}^{6} x p(x)$$

$$= \frac{1}{10}(-1) + \frac{3}{10}(1) + \frac{1}{10}(1) + \frac{2}{10}(2) = \frac{7}{10}$$

$$E[z] = \sum_{1}^{3} z p(z)$$

$$= \left(\frac{1}{10} + \frac{2}{10}\right)(-1) + \left(\frac{1}{10} + \frac{2}{10}\right)(1) = 0$$

$$\mathrm{var}[x] = \sum_{1}^{6} (x - \bar{x})^2 p(x)$$

$$= \frac{1}{10}\left(\frac{17}{10}\right)^2 + \frac{2}{10}\left(\frac{7}{10}\right)^2 + \frac{1}{10}\left(\frac{7}{10}\right)^2 + \frac{3}{10}\left(\frac{3}{10}\right)^2$$

$$+ \frac{1}{10}\left(\frac{3}{10}\right)^2 + \frac{2}{10}\left(\frac{13}{10}\right)^2 = \frac{81}{100}$$

$$\mathrm{var}[z] = \sum_{1}^{3} (z - \bar{z})^2 p(z)$$

$$= \left(\frac{1}{10} + \frac{2}{10}\right)(-1)^2 + \left(\frac{1}{10} + \frac{2}{10}\right)(1)^2 = \frac{3}{5}$$

$$\text{cov}[x,z]=\sum_{1}^{3}(x-\overline{x})(z-\overline{z})p(x,z)$$

$$=\frac{1}{10}\left(-\frac{17}{10}\right)(-1)+\frac{2}{10}\left(-\frac{7}{10}\right)(-1)$$

$$+\frac{1}{10}\left(\frac{3}{10}\right)+\left(\frac{2}{10}\right)\left(\frac{13}{10}\right)=\frac{3}{5}$$

代入线性最小方差估计公式(3-4-11)和(3-4-12)中得

$$\hat{x}_{lmv}=Ex+\text{cov}(x,z)(\text{var}\,z)^{-1}(z-Ez)$$

$$=\frac{7}{10}+\frac{3}{5}\times\frac{5}{3}(z-0)$$

$$=\frac{7}{10}+z$$

$$=\begin{cases}-\dfrac{3}{10}, & z=-1\\ \dfrac{7}{10}, & z=0\\ \dfrac{17}{10}, & z=1\end{cases}$$

$$P_{lmv}=E(x-\hat{x}_{lmv})^2=\text{var}\,x-\text{cov}(x,z)(\text{var}\,z)^{-1}\text{cov}(zx)$$

$$=\frac{81}{100}-\frac{3}{5}\times\frac{5}{3}\times\frac{3}{5}=\frac{21}{100}$$

或 $$P_{lmv}=\sum_{1}^{6}(x-\hat{x}_{lmv})^2P(x)$$

$$=\frac{1}{10}\left(\frac{7}{10}\right)^2+\frac{2}{10}\left(\frac{3}{10}\right)^2+\frac{1}{10}\left(\frac{17}{10}\right)^2+\frac{3}{10}\left(\frac{3}{10}\right)^2+\frac{1}{10}\left(\frac{7}{10}\right)^2+\frac{2}{10}\left(\frac{3}{10}\right)^2=\frac{21}{100}$$

下面再求最小方差估计$\hat{x}_{mv}$。

已知 $$\hat{x}_{mv}=E[x/z]=\frac{\sum_{1}^{2}xp(x/z)}{\sum_{1}^{2}p(x/z)}$$

把已知条件代入，得

$$z=-1,\quad \hat{x}_{mv}=\frac{-1\times\frac{1}{10}+0\times\frac{2}{10}}{\frac{1}{10}+\frac{2}{10}}=-\frac{1}{3}$$

$$z=0,\quad \hat{x}_{mv}=\frac{0\times\frac{1}{10}+1\times\frac{3}{10}}{\frac{1}{10}+\frac{3}{10}}=\frac{3}{4}$$

$$z=1,\quad \hat{x}_{mv}=\frac{1\times\frac{1}{10}+2\times\frac{2}{10}}{\frac{1}{10}+\frac{2}{10}}=\frac{5}{3}$$

又 $$P_{mv}=E[(x-\hat{x}_{mv})^2]=\sum_{i}^{t}(x-\hat{x}_{mv})^2P(x/z)$$

$$=\frac{1}{10}\left(\frac{2}{3}\right)^2+\frac{2}{10}\left(\frac{1}{3}\right)^2+\frac{1}{10}\left(\frac{3}{4}\right)^2+\frac{3}{10}\left(\frac{1}{4}\right)^2$$

$$+\frac{1}{10}\left(\frac{2}{3}\right)^2+\frac{2}{10}\left(\frac{1}{3}\right)^2=\frac{5}{24}$$

可以看到，$P_{mv}<P_{lmv}$。也就是说，最小方差估计质量要优于线性最小方差估计。$\hat{x}_{mv}$ 和 $\hat{x}_{lmv}$ 根据以前的分析，它们都是 x 的无偏估计，但估计值不必相同。

线性最小方差估计的递推算法

在根据 N 个测量值得出了线性最小方差估计 $\hat{X}_{lmv}$ 及其方差阵 P_{lmv} 之后，如果再增加一个测量 z_{N+1}，重算一遍，计算量太大了，若采用递推算法就方便多了。

由例 3-4-1 得 N 个测量值的线性最小方差估计公式为

$$\left.\begin{aligned}&K_N=PH_N^T(H_NPH_N^T+R_N)^{-1}\\&\hat{X}_N=\mu+K_N(Z_N+H_N\mu)\\&P_N=(I-K_NH_N)P\end{aligned}\right\}\tag{3-4-23}$$

其中 μ、P 为 X 的先验均值和方差。

利用附录D中矩阵等式，可把上面结果加以改写，得

$$\hat{X}_N=P_NP^{-1}\mu+P_NH_N^TR_N^{-1}Z_N\tag{3-4-24}$$

$$P_N^{-1}=P^{-1}+H_N^TR_N^{-1}H_N\tag{3-4-25}$$

现在加入了新测量值

$$z_{N+1}=h_{N+1}X+e_{N+1}\tag{3-4-26}$$

其中噪声 $E[e_{N+1}]=0$，$E[e_{N+1}e_{N+1}^T]=r_{N+1}$。

记 $$Z_{N+1}=\begin{pmatrix}Z_N\\z_{N+1}\end{pmatrix}\qquad H_{N+1}=\begin{pmatrix}H_N\\h_{N+1}\end{pmatrix}$$

$$R_{N+1}=\begin{pmatrix}R_N&0\\0&r_{N+1}\end{pmatrix}$$

则得到新估计为

$$\hat{X}_{N+1}=P_{N+1}P^{-1}\mu+P_{N+1}H_{N+1}^TR_{N+1}^{-1}Z_{N+1}\tag{3-4-27}$$

$$P_{N+1}^{-1}=P^{-1}+H_{N+1}^TR_{N+1}^{-1}H_{N+1}\tag{3-4-28}$$

由(3-4-28)式，得

$$P_{N+1}^{-1} = P^{-1} + [H_N^T \quad h_{N+1}^T] \begin{pmatrix} R_N^{-1} & 0 \\ 0 & r_{N+1}^{-1} \end{pmatrix} \begin{pmatrix} Z_N \\ \vdots \\ z_{N+1} \end{pmatrix}$$

$$= P^{-1} + H_N^T R_N^{-1} H_N + h_{N+1}^T r_{N+1}^{-1} h_{N+1}$$

$$= P_N^{-1} + h_{N+1}^T r_{N+1}^{-1} h_{N+1} \qquad (3\text{-}4\text{-}29)$$

由(3-4-27)式，得

$$P_{N+1}^{-1} \hat{X}_{N+1} = P^{-1}\mu + [H_N^T \quad h_{N+1}^T] \begin{pmatrix} R_N^{-1} & 0 \\ 0 & r_{N+1}^{-1} \end{pmatrix} \begin{pmatrix} Z_N \\ z_{N+1} \end{pmatrix}$$

$$= P^{-1}\mu + [H_N^T R_N^{-1} H_N + h_{N+1}^T r_{N+1}^{-1} h_{N+1}$$

$$= P_N^{-1} \hat{X} + h_{N+1}^T r_{N+1}^{-1} h_{N+1}$$

$$\hat{X}_{N+1} = P_{N+1}[P_N^{-1} \hat{X}_N + h_{N+1}^T r_{N+1}^{-1} h_{N+1}] \qquad (3\text{-}4\text{-}30)$$

由(3-4-29)、(3-4-30)式，再利用附录D中矩阵等式，可得到新估计的递推算法为

$$K_{N+1} = P_N h_{N+1}^T [h_{N+1} P_N h_{N+1}^T + r_{N+1}]^{-1} \qquad (3\text{-}4\text{-}31)$$

$$\hat{X}_{N+1} = \hat{X}_N + K_{N+1}[z_{N+1} - h_{N+1} \hat{X}_N] \qquad (3\text{-}4\text{-}32)$$

$$P_{N+1} = [I - K_{N+1} h_{N+1}] P_N \qquad (3\text{-}4\text{-}33)$$

注意到上述递推算法与最优加权最小二乘估计的递推算法相类似。

§3-5 最小均方误差估计$\hat{X}_{lmse}$

在求线性最小方差估计$\hat{X}_{lmv}$时，我们不需要知道完整的概率分布知识，但需要知道X和Z的头两阶矩，有时这一点也不具备，所以我们希望找到一种要求更少先验统计知识的估计方法，这种估计方法就是最小均方误差估计$\hat{X}_{lmse}$，又称为正交估计。得出$\hat{X}_{lmse}$时所根据的准则是最小均方误差准则。下面就来研究求最小均方误差估计的方法。

规定估计$\hat{X}$是观测集合的加权和

$$\hat{X} = \sum_{i=1}^{N} a_i z_i = A^T Z \qquad (3\text{-}5\text{-}1)$$

其中 $A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}$， $Z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix}$

估计$\hat{X}$的均方误差为

$$J = \mathrm{tr} E[(X - A^T Z)(X - A^T Z)^T]$$

$$= E[(X - A^T Z)^T (X - A^T Z)] \qquad (3\text{-}5\text{-}2)$$

若x为标量，则均方误差为

$$J = E[(x - A^T Z)^2]$$

选择A使J达到最小，此时的A称为最佳权系数，并得到最小均方误差估计$\hat{X}_{lmse}$。令(3-5-2)式对A^T求导，并使之等于零，

$$\frac{\partial J}{\partial A^T}=\frac{\partial}{\partial A^T}E[(X-A^TZ)^T(X-A^TZ)]$$

$$=-2E[(X-A^TZ)Z^T]=0 \tag{3-5-4}$$

解出 $$E[XZ^T]=E[A^TZZ^T]=A^TE[ZZ^T] \tag{3-5-5}$$

得最佳权系数矩阵

$$A^T_{lmse}=E[XZ^T](E[ZZ^T])^{-1} \tag{3-5-6}$$

或 $$A_{lmse}=(E[ZZ^T])^{-1}E[ZX^T] \tag{3-5-7}$$

因为$E[ZZ^T]$是测量序列的自相关矩阵，$E[XZ^T]$为测量序列Z与待估计量X的互相关矩阵。记 $$\Phi_{zz}=E[ZZ^T]$$

$$\Phi_{zx}=E[ZX^T]$$

则有 $$A_{lmse}=\Phi_{zz}^{-1}\Phi_{zx} \qquad A^T_{lmse}=\Phi^T_{zx}\Phi^{-1}_{zz} \tag{3-5-8}$$

A_{lmse}是为了得到最小均方误差估计的最佳权矩阵，于是，得到最小均方误差估计为

$$\hat{X}_{lmse}=A^T_{lmse}Z=\Phi^T_{zx}\Phi^{-1}_{zz}Z \tag{3-5-9}$$

相应的最小均方误差阵为

$$P_{lmse}=E[(X-\hat{X}_{lmse})(X-\hat{X}_{lmse})^T]$$

$$=E[(X-\hat{X}_{lmse})X^T]-E[(X-\hat{X}_{lmse})Z^TA_{lmse}] \tag{3-5-10}$$

由(3-5-4)式得知，上式右边第二项等于零，故

$$P_{lmse}=E[XX^T]-E[A^T_{lmse}ZX^T]$$

$$=\Phi_{xx}-\Phi^T_{zx}\Phi^{-1}_{zz}\Phi_{zx} \tag{3-5-11}$$

其中Φ_{xx}是待估计信号的自相关矩阵。若x为标量，则$\Phi_{xx}=E[x^2]$，即为信号均方值。

(3-5-9)和(3-5-11)为最小均方误差估计及其均方误差的表达式，必要的先验知识是测量序列的自相关矩阵Φ_{zz}，测量序列Z与待估计量X的互相关矩阵Φ_{zx}，以及X的自相关矩阵Φ_{xx}。

当Z与X均为平稳随机序列时，Φ_{zz}、Φ_{zx}和Φ_{xx}是容易求出的，而当Z与X为非平稳随机序列时，一般说来是相当困难的。把最小均方误差估计准则应用于平稳随机时间过程就构成了维纳滤波理论的基础，关于这个问题留在下章中详细讨论。下面一个简单例子说明求最小均方误差估计的方法。

例3-5-1 设有线性矢量测量方程

$$Z=HX+e \tag{3-5-12}$$

已知 $E[XX^T]=\Phi_{xx}=S$，$E[e]=0$，$E[ee^T]=R$，

$E[Xe^T]=0$

求最小均方误差估计$\hat{X}_{lmse}$及其均方误差矩阵。

解：根据已知条件可以求出

$$\Phi_{zz}=E[(HX+e)(HX+e)^T]$$

$$=HSH^T+R \tag{3-5-13}$$

$$\Phi_{zx}=E[(HX+e)X^T)]=HS \tag{3-5-14}$$

代入(3-5-9)和(3-5-11)式，得

$$\hat{X}_{lmse}=\Phi^T_{zx}\Phi^{-1}_{zz}Z$$

$$= SH^T(HSH^T + R)^{-1}Z \tag{3-5-15}$$

$$= (S^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} Z \tag{3-5-16}$$

$$P_{lmse} = \Phi_{xx} - \Phi_{zx}^T \Phi_{zz}^{-1} \Phi_{zx}$$

$$= S - SH^T(HSH^T + R)^{-1}HS \tag{3-5-17}$$

$$= (S^{-1} + H^T R^{-1} H)^{-1} \tag{3-5-18}$$

一种常见的情况是：x 是标量随机变量，测量是多次的，测量方程为

$$z_i = x_i + e_i, \quad i = 1, 2, \cdots, N$$

已知 $E[x^2] = S$，$E[e_i] = 0$，$E[e_i^2] = \sigma_e^2$，各e_i互相独立。

由(3-5-16)与(3-5-18)式容易求得

$$\hat{X} = \frac{1}{N + \frac{\sigma_e^2}{S}} \sum_{i=1}^{N} z_i \tag{3-5-19}$$

$$P_{lmse} = \frac{\sigma_e^2}{N + \frac{\sigma_e^2}{S}} \tag{3-5-20}$$

倘若各噪声的方差不相等，$E[e_i^2] = \sigma_i^2$，则求得

$$\hat{X}_{lmse} = \left(\frac{1}{\frac{1}{S} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \right) \sum_{i=1}^{N} \frac{z_i}{\sigma_i^2} \tag{3-5-21}$$

$$P_{lmse} = \frac{1}{\frac{1}{S} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \tag{3-5-22}$$

现在把 $\hat{X}_{lmse}$ 写成测量值 z_i 的加权和的形式

$$\hat{X}_{lmse} = \sum_{i=1}^{N} a_i z_i \tag{3-5-23}$$

对于(3-5-19)式，各权系数 a_i 是相等的。

$$a_1 = a_2 = \cdots = a_N = \frac{1}{N + \frac{\sigma_e^2}{S}} \tag{3-5-24}$$

对于(3-5-21)式而言，各权系数是不相等的。

$$a_i = \frac{\frac{1}{\sigma_i^2}}{\frac{1}{S} + \sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \qquad i = 1, 2, \cdots, N \tag{3-5-25}$$

下再面对最小均方误差估计的性质作几点说明：

1. 偏量

偏量 $$b = E[x - \hat{x}_{lmse}]$$

$$= E[x] - \Phi_{zx}^T \Phi_{zz}^{-1} E[Z] \neq 0 \tag{3-5-26}$$

所以，$\hat{X}_{lmse}$ 是一种有偏估计。

验证(3-5-26)式，由(3-5-19)式得

偏量
$$b = Ex - \frac{1}{N + \frac{\sigma_v^2}{S}} E\left[\sum_{i=1}^{N} z_i\right]$$

$$= \overline{x} - \frac{N\overline{x}}{N + \frac{\sigma_v^2}{S}} = \frac{\overline{x}}{1 + N\frac{S}{\sigma_v^2}} \tag{3-5-27}$$

但是当测量次数无限增多时，有

$$\lim_{N\to\infty} b = \lim_{N\to\infty} \frac{\overline{x}}{1 + N\frac{S}{\sigma_v^2}} = 0 \tag{3-5-28}$$

不失一般性，可以得知，$\hat{x}_{lmse}$ 是一种渐近无偏估计。

2. 方差

由(3-5-20)式，可得

$$\lim_{N\to\infty} P_{lmse} = \lim_{N\to\infty} \frac{\sigma_v^2}{N + \frac{\sigma_v^2}{S}} = 0 \tag{3-5-29}$$

不失一般性，可以得知，$\hat{x}_{lmse}$ 是一致性估计，随着测量次数增多，偏量和方差都趋于零。

3. 正交性

由(3-5-4)式得
$$E[(X - \hat{X}_{lmse})^T Z] = 0 \tag{3-5-30}$$

从矢量空间意义上看，意味着误差矢量 $X - \hat{X}_{lmse}$ 与测量矢量Z正交。设想由测量集$\{Z_i\}$构成超平面，而矢量X不在此超平面上，既然估计误差（$X - \hat{X}_{lmse}$)与此超平面正交，那末，估计$\hat{X}_{lmse}$就是X在此超平面上的投影，其关系可以用图3-5-1中的几何图形表示。

上述投影关系称为正交原理。因为$\hat{X}_{lmse}$可由正交原理导出，所以最小均方误差估计又称为正交估计。

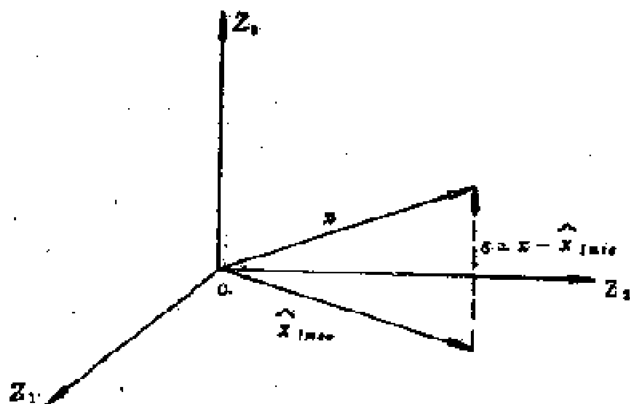


图3-5-1 正交原理的几何解释

§3-6 变量估计方法总结

在第二章和这一章中，已研究了许多种估计方法，有必要在这里进行比较和总结。

1. 最小二乘估计$\hat{X}_{ls}$

在最小二乘估计方法中，待估计量认为是确定量，不需要任何先验统计知识，依据的

准则是估计误差的平方和达到最小。当对噪声的先验统计知识缺乏时，这是一种十分有用的估计方法。

2．最优加权最小二乘估计$\hat{X}_{lsw}$

待估计量同样认为是确定量，依据的准则是估计误差的加权平方和达到最小，但要求测量噪声的协方差阵为已知。若各测量噪声独立，且方差相等，则$\hat{X}_{lsw}$与$\hat{X}_{ls}$相同。

3．最大似然估计$\hat{X}_{ml}$

待估计量认为是确定量，要求知道条件概率密度$p(Z/X)$的完整知识，寻求与$p(Z/X)$最大值对应的值作为估计量，在多数情况下，其估计误差的方差不容易求得，但容易求得方差的下界，此方差下界称为克拉美-罗下界。在高斯分布条件下，最大似然估计等于最优加权最小二乘估计。

4．最大后验估计$\hat{X}_{map}$

待估计量是随机变量，要求知道后验概率密度$p(X/Z)$的完整知识，寻求$p(X/Z)$最大值对应的X值作为估计值，其方差一般难以求得，但可求得其方差下界。

最大似然估计可以看成是最大后验估计的一个特例，最大似然估计是被估计量为非随机量这一特殊情况下的最大后验估计。

5．最小方差估计$\hat{X}_{mv}$

待估计量是随机变量，要求知道后验密度$p(X/Z)$的完整知识，$p(X/Z)$的均值就是最小方差估计。$\hat{X}_{mv}$是一切估计中方差最小的无偏估计，所以它在理论上是最优估计。

最大似然估计、最大后验估计、最小方差估计总称为贝叶斯估计。

6．线性最小方差估计$\hat{X}_{lmv}$

待估计量是随机变量，规定估计值是测量值的线性函数，要求知道X和Z的头两阶矩，依据的准则是均方误差达到最小。$\hat{X}_{lmv}$是一切线性估计中方差最小的无偏估计。

7．最小均方误差估计$\hat{X}_{lmse}$

待估计量是随机变量，规定估计值是测量值的加权组合，要求的先验知识是测量集合的自相关矩阵和测量集合与待估计变量的互相关矢量（或矩阵），依据的准则是均方误差达到最小。最小均方误差估计是渐近无偏的一致性估计。对于零均值序列，显见最小均方误差估计与线性最小方差估计相同。

在遇到某具体工程问题时，究竟选用什么样的估计方法，要根据具体情况而定。一般说来，精度高的估计方法需要的先验统计知识较多；精度低的估计方法需要的先验统计知识较少。下面举几个工程上的例子加以说明。

例3-6-1 这是一个雷达工程上遇到的问题。设雷达发射一确知信号序列s_i，$i=1, 2, \cdots, N$，接收序列是经过随机信道调制并附加有可加性白噪声，接收信号为，$z_i=x_is_i+e_i$，$i=1, 2, \cdots, N$。求对振幅调制因子x的估计。

解：此问题的解法可以有多种，根据对信道先验知识的多少，分不同情况采用不同的估计方法。

1．对信道没有任何先验统计知识，这时可把x看成是确定量，采用最小二乘估计方法。

把测量序列写成矢量形式

$$Z=Hx+e \tag{3-6-1}$$

其中 $Z=\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{pmatrix}$, $H=\begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{pmatrix}$, $e=\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{pmatrix}$

根据最小二乘估计表达式，得

$$\hat{X}_{ls}=(H^TH)^{-1}H^TZ$$

$$=\frac{\sum_{i=1}^{N} s_i z_i}{\sum_{i=1}^{N} s_i^2} \tag{3-6-2}$$

分子$\sum_{i=1}^{N} s_i z_i$等价于脉冲列匹配滤波器的输出，分母$\sum_{i=1}^{N} s_i^2$等价已知发射脉冲列的能量，所以对振幅调制因子x的估计可以用图3-6-1所示之结构实现。

2．假定雷达发射的是等幅脉冲列，$s_i=s$，已知接收功率信噪比$\frac{s^2E[x^2]}{\sigma_e^2}$的先验知识，这时可采用最小均方误差估计方法。

$$\hat{x}_{lmse}=\Phi_{xz}^T\Phi_{zz}^{-1}Z \tag{3-6-3}$$

图3-6-1　实现(3-6-2)式之结构，求$\hat{x}_{ls}$

记　　$p_x=E[x^2]$

则　　$$\hat{x}_{lmse}=(P_x^{-1}+H^TR^{-1}H)^{-1}H^TR^{-1}Z$$

$$=\frac{1}{s\left(N+\frac{\sigma_e^2}{s^2p_x}\right)}\sum_{i=1}^{N} z_i \tag{3-6-4}$$

把$\hat{x}_{lmse}$写成测量序列加权和的形式，

令　　$$a_i=\frac{1}{s\left(N+\frac{\sigma_e^2}{s^2p_x}\right)} \tag{3-6-5}$$

则有　　$$\hat{x}_{lmse}=\sum_{i=1}^{N} a_i z_i \tag{3-6-6}$$

所以，估计$\hat{x}_{lmse}$可以用图3-6-2所示之横向滤波器实现，滤波器的权为$a_1=a_2=\cdots a_N=\frac{1}{s\left(N+\frac{\sigma_e^2}{s^2p_x}\right)}$。

3．已知x，及噪声e_i均为高斯分布，且统计参数已知，这时我们已具有完整的概率分布知识，因此，可以求x的最大后验估计$\hat{x}_{map}$。

已知最大后验估计方程是

$$\left.\frac{\partial}{\partial x}\ln p(Z/x)+\frac{\partial}{\partial x}\ln p(x)\right|_{x=\hat{x}_{map}}=0 \tag{3-6-7}$$

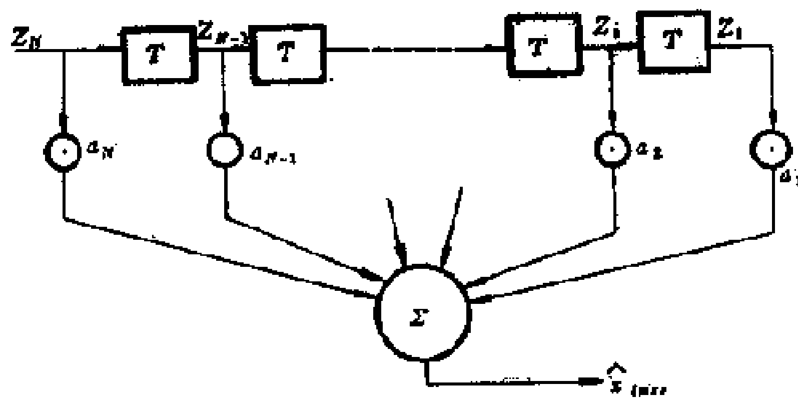


图3-6-2 用横向滤波器求$\hat{x}_{lms}$

由已知条件得出

$$p(x)=\frac{1}{\sigma_x\sqrt{2\pi}}\exp\left[-\frac{(x-\bar{x})^2}{2\sigma_x^2}\right] \tag{3-6-8}$$

$$p(z/x)=\prod_{i=1}^{N}\frac{1}{\sigma_v\sqrt{2\pi}}\exp\left[-\frac{(z_i-s_ix)^2}{2\sigma_v^2}\right] \tag{3-6-9}$$

代入(3-6-7)式，得

$$-\frac{x_{map}-\bar{x}}{\sigma_x^2}+\frac{1}{\sigma_v^2}\sum_{i=1}^{N}s_i(z_i-s_i\hat{x}_{map})=0 \tag{3-6-10}$$

解得

$$\hat{x}_{map}=\frac{\bar{x}+\dfrac{\sigma_x^2}{\sigma_v^2}\sum\limits_{i=1}^{N}s_iz_i}{1+\dfrac{\sigma_x^2}{\sigma_v^2}\sum\limits_{i=1}^{N}s_i^2} \tag{3-6-11}$$

当$\bar{x}=0$和$\sigma_x^2\gg\sigma_v^2$时，得

$$\hat{x}_{map}\approx\frac{\sum\limits_{i=1}^{N}s_iz_i}{\sum\limits_{i=1}^{N}s_i^2} \tag{3-6-12}$$

也就是说，在上述条件下，最大后验估计近似等于最小二乘估计。

4．若$p(x)$为瑞利分布，其他条件不变，仍然可以求x的最大后验估计。

由
$$p(x)=\frac{x}{\sigma_x^2}e^{-\frac{x^2}{2\sigma_x^2}} \qquad x\geqslant 0 \tag{3-6-13}$$

$$p(x)=0 \qquad x<0$$

得
$$\frac{\partial}{\partial x}\ln p(x)=\frac{1}{x}-\frac{x}{\sigma_x^2} \tag{3-6-14}$$

代入最大后验估计方程，得

$$\hat{x}_{map}=\frac{y}{2R}\left(1+\sqrt{\frac{4R}{y^2}+1}\right) \tag{3-6-15}$$

其中 $$y=\frac{1}{\sigma_e^2}\sum_{i=1}^{N} s_i z_i \tag{3-6-16}$$

$$R=\frac{1}{\sigma_x^2}+\frac{1}{\sigma_e^2}\sum_{i=1}^{N} s_i^2 \tag{3-6-17}$$

显然，R 为一已知量，y 代表脉冲列匹配滤波器的输出，对 y 进行非线性运算才可解出 $\hat{x}_{map}$。求 $\hat{x}_{map}$ 可用图 3-6-3 所示之结构实现。

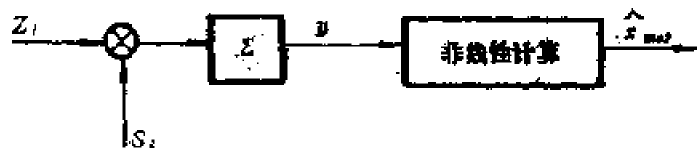


图3-6-3 实现(3-6-15)式之结构，求$\hat{x}_{map}$。

对信号能量进行归一化，令 $\sum_{i=1}^{N} s_i^2=1$，则有

$$\frac{y_2}{4R}=\frac{\left(\sum_{i=1}^{N}\frac{s_i z_i}{2\sigma_e}\right)^2}{1+\frac{\sigma_e^2}{\sigma_x^2}} \tag{3-6-18}$$

分两种极端情况加以讨论。

a. 高信噪比情况，$\frac{\sigma_e^2}{\sigma_x^2}\ll 1$

这时 $\frac{4R}{y^2}\ll 1$，因此得

$$\hat{x}_{map}\approx\frac{y}{R}\approx\sum_{i=1}^{N} s_i z_i \tag{3-6-19}$$

这时估计器就是一个脉冲列匹配滤波器。

b. 小信噪比情况，$\frac{\sigma_e^2}{\sigma_x^2}\gg 1$

这时 $\frac{4R}{y^2}\gg 1$

$$\hat{x}_{map}\approx\frac{1}{\sqrt{R}}\approx\sigma_x \tag{3-6-20}$$

显见，估计 $\hat{x}_{map}$ 与测量值无关，也就是说，我们不能从测量值得到 x 的合理估计。

若已知第 i 个发射脉冲的振幅 s_i，则通过信道之后的振幅的估计是 $\hat{x}s_i$。

综上所述，对于同一信号模型，由于采用估计方法的不同，得到的估计也不同。

例3-6-2 连续时间信号振幅的测量

设接收波形具有如下形式

$$z(t)=As(t)+n(t)\qquad 0\leqslant t\leqslant t_f \tag{3-6-21}$$

其中$s(t)$具有已知形式。$n(t)$是高斯白噪声，其功率谱密度为N_0，求振幅A。

解：根据题意，我们对振幅A没有任何先验概率知识。把A作为确定量看待，因而可采用最大似然估计方法求$\hat{A}_{ml}$。

因为(3-6-21)式是连续形式，为了应用离散估计方法，首先对信号进行离散采样，然后令样本加密，并把各样本写成矢量形式，得

$$\begin{pmatrix} z(T) \\ z(2T) \\ \vdots \\ z(NT) \end{pmatrix} = \begin{pmatrix} s(T) \\ s(2T) \\ \vdots \\ s(NT) \end{pmatrix} A + \begin{pmatrix} n(T) \\ n(2T) \\ \vdots \\ n(NT) \end{pmatrix} \tag{3-6-22}$$

已知测量噪声是白噪声，故有

$$p(Z/x) = p(z(T)/x)\cdot p(z(2T)/x)\cdots p(z(NT)/x) \tag{3-6-23}$$

于是我们可以写出条件密度函数

$$p(Z/A) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(z_i - As_i)^2\right] \tag{3-6-24}$$

令$N\to\infty$，$T\to 0$，则有

$$\sum_{i=1}^{N}(z_i - As_i)^2 \to \frac{1}{t_f}\int_0^{t_f}[z(t) - As(t)]^2 dt$$

又 $\sigma^2 = N_0 B = \frac{N_0}{2t_f}$，$t_f$为接收信号持续时间。

于是，接收波形$z(t)$的条件概率密度函数为

$$p(z(t)/A) = k\exp\left\{-\frac{1}{N_0}\int_0^{t_f}[z(t) - As(t)]^2 dt\right\} \tag{3-6-25}$$

其中k是常数。因此，最大似然估计方程为

$$\int_0^T [z(t) - As(t)]\frac{\partial As(t)}{\partial A}\Big|_{A=\hat{A}_{ml}} dt = 0 \tag{3-6-26}$$

其等价形式为

$$\int_0^T [z(t) - \hat{A}_{ml}s(t)]s(t)\,dt = 0 \tag{3-6-27}$$

故信号振幅A的最大似然估计为

$$\hat{A}_{ml} = \frac{\int_0^T z(t)s(t)\,dt}{\int_0^T s^2(t)\,dt} \tag{3-6-28}$$

其中分母表示$s(t)$的能量，是已知值，分子表示相关器或匹配滤波器的输出。求A_{ml}可用图3-6-4所示之结构实现。

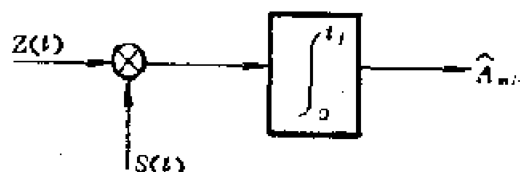


图3-6-4 实现(3-6-28)式之结构，求$\hat{A}_{ml}$

对信号能量进行归一化，令$\int_0^T s^2(t)dt=1$，则有

$$\hat{A}_{ml}=\int_0^{t_f} z(t)s(t)dt \tag{3-6-29}$$

$\hat{A}_{ml}$ 的均值为

$$E[\hat{A}_{ml}]=E\left[\int_0^{t_f}(As(t)+n(t))s(t)dt\right]=A \tag{3-6-30}$$

所以$\hat{A}_{ml}$是振幅A的无偏估计。

由克拉美-罗不等式得

$$\mathrm{var}(\hat{A}-A)\geqslant\frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{Z}/A)}{\partial A^2}\right]}$$

得
$$\frac{\partial}{\partial A}\ln p(\mathbf{Z}/A)=\frac{2}{N_0}\int_0^{t_f}[z(t)-As(t)]s(t)dt \tag{3-6-31}$$

$$\frac{\partial^2}{\partial A^2}\ln p(\mathbf{Z}/A)=-\frac{2}{N_0}\int_0^{t_f} s^2(t)dt=-\frac{2}{N_0} \tag{3-6-32}$$

由（3-6-31）式可判断$\hat{A}_{ml}$是振幅A的优效估计。于是求得估计误差方差等于克拉美-罗下界，有

$$\mathrm{var}[\hat{A}_{ml}-A]=\frac{N_0}{2} \tag{3-6-33}$$

例3-6-3　连续时间信号相位的测量

设接收波形为

$$z(t)=A\cos(\omega_c t-\varphi)+n(t)\qquad 0\leqslant t\leqslant t_f \tag{3-6-34}$$

其中振幅A和频率ω_0为已知，要测量的是初相位φ。$n(t)$为高斯白噪声，功率谱密度为N_0，求初相位φ的估计。

解：由于对φ没有任何先验概率知识，所以求它的最大似然估计$\hat{\varphi}_{ml}$。

类似于上例，先对$z(t)$采样，再令样本变密，写出接收波形$z(t)$的条件密度

$$p(z(t)/\varphi)=k\exp\left[-\frac{1}{N_0}\int_0^{t_f} z(t)-s(t,\varphi)\right]^2 dt \tag{3-6-35}$$

其中k为常数。

代入最大似然估计方程

$$\left.\frac{\partial}{\partial\varphi}\ln p(z(t)/\varphi)\right|_{\varphi=\hat{\varphi}_{ml}}=0$$

得
$$\left.\int_0^{t_f}[z(t)-s(t,\varphi)]\frac{\partial s(t,\varphi)}{\partial\varphi}dt\right|_{\varphi=\hat{\varphi}_{ml}}=0 \tag{3-6-36}$$

代入$s(t,\varphi)=A\cos(\omega_0 t-\varphi)$，得

$$\int_0^{t_f}[z(t)-A\cos(\omega_0 t-\varphi)]\sin(\omega_0 t-\varphi)dt=0 \tag{3-6-37}$$

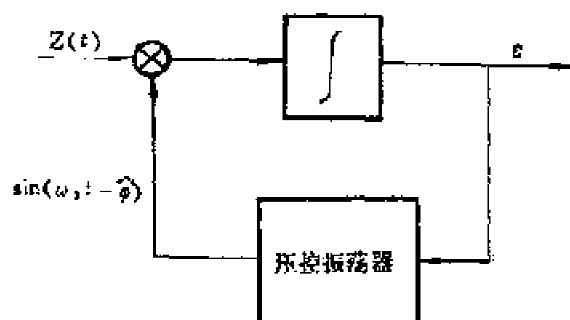


图3-6-5 实现(3-6-38)式之锁相环结构，求$\hat{\varphi}_{ml}$

若$\omega_0 t_f \gg 1$，则此积分中的第二项可以略去，故相位估计$\hat{\varphi}_{ml}$是下列方程的解

$$\int_0^{t_f} z(t)\sin(\omega_0 t - \hat{\varphi}_{ml})dt = 0 \qquad (3\text{-}6\text{-}38)$$

实现(3-6-38)式之计算可用图3-6-5所示之锁相环结构。

此外还可以用双通道正交系统来实现相位估计，把$p(z(t)/\varphi)$写成如下形式

$$p(z(t)/\varphi) = k\cdot\exp\left\{-\frac{E_T}{N_0}\right\}\cdot\exp\left\{\frac{2}{N_0}\int_0^{t_f} z(t)s(t\varphi,)dt\right\}$$

$$= k\exp\left\{-\frac{E_T}{N_0}\right\}\cdot\exp\left\{\frac{2}{N_0}\xi_\varphi\right\} \qquad (3\text{-}6\text{-}39)$$

其中 $$E_T = \int_0^{t_f} s^2(t,\varphi)dt \qquad (3\text{-}6\text{-}40)$$

$$\xi_\varphi = \int_0^{t_f} z(t)s(t,\varphi)dt \qquad (3\text{-}6\text{-}41)$$

这里，ξ_φ是相关积分，使ξ_φ达到最大的φ值也就是使$p(z(t)/\varphi)$达到最大的φ值，因此可以从相关积分出发来找求相位φ的最大似然估计$\hat{\varphi}$。

由于信号

$$s(t,\varphi) = A\cos(\omega_0 t - \varphi)$$
$$= A\cos\varphi\cos\omega_0 t + A\sin\varphi\sin\omega_0 t$$

故有

$$\xi_\varphi = \int_0^{t_f} z(t)A\cos(\omega_0 t - \varphi)dt$$
$$= \cos\varphi\int_0^{t_f} Az(t)\cos\omega_0 t\,dt$$
$$+ \sin\varphi\int_0^{t_f} Az(t)\sin\omega_0 t\,dt \qquad (3\text{-}6\text{-}42)$$
$$= \xi_c\cos\varphi + \xi_s\cos\varphi$$

其中 $$\xi_c = \int_0^{t_f} z(t)A\cos\omega_0 t\,dt$$

$$\xi_s = \int_0^{t_f} z(t)A\sin\omega_0 t\,dt$$

引进新变量ρ和θ，进行变量变换，使得

$$\rho = \sqrt{\xi_c^2 + \xi_s^2}, \quad \cos\theta = \frac{\xi_c}{\rho}, \quad \sin\theta = \frac{\xi_s}{\rho} \qquad (3\text{-}6\text{-}43)$$

这样，相关积分成为

$$\xi_\varphi = \rho\cos(\theta - \varphi) \qquad (3\text{-}6\text{-}44)$$

其中φ为未知初相位，θ是一个可由ξ_c和ξ_s来计算的量，当$\theta=\varphi$时，$\cos(\theta-\varphi)=1$，则ξ_φ达到最大值，所以，未知初相位φ的估计为

$$\hat{\varphi}_{ml} = \hat{\theta} = \tan^{-1}\frac{\xi_s}{\xi_c}$$

$$= \tan^{-1}\frac{\int_0^{t_f} z(t)\sin\omega_0 t\,dt}{\int_0^{t_f} z(t)\cos\omega_0 t\,dt} \tag{3-6-45}$$

实现（3-6-45）式之计算可用如图 3-6-6 所示之结构。顺便指出，因为比值 ξ_s/ξ_c 与振幅无关，所以，即使振幅 A 未知，也不会影响对相位 φ 的估计。

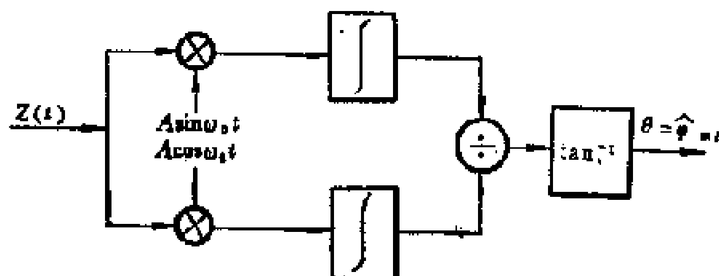


图3-6-6　实现(3-6-45)式之结构，求 $\hat{\varphi}_{ml}$

第二篇 维纳滤波

第四章 离散维纳滤波

前两章中研究了变量的估计问题，包括确定变量估计和随机变量估计，从本章起将研究随机时间序列的估计问题。

设测量值是一随机时间序列，由信号x_i和噪声e_i混合而成。

$$z_i = x_i + e_i \qquad i = 1, 2, 3, \cdots$$

其中信号x_i本身也是随机的，我们希望设计一个滤波器，对测量序列z_i进行过滤，使得在滤波器的输出端能较好地复现信号x_i或它的线性函数，把期望复现的信号称为期望信号，滤波器的设计目的就是希望得到对期望信号的合理估计。

离散维纳滤波是建立在横向滤波器基础之上的，横向滤波器的权按照最小均方误差准则调整到最佳，就构成了维纳滤波器，所以维纳滤波是上章中最小均方误差估计的实现。

本章中除了研究经典的维纳滤波理论之外，还加入了近年来的新进展，如权系数的递推算法，广义维纳滤波等。对于近十余年来新发展起来的自适应维纳滤波将留在下章中研究。最后还研究了连续时间信号的维纳滤波器。

§4-1 平稳随机序列维纳滤波

设有测量序列

$$z_i = x_i + e_i \qquad i = 1, 2, 3, \cdots, N \tag{4-1-1}$$

其中x_i为平稳信号序列，e_i为平稳噪声序列，平稳序列之和仍为平稳序列，故z_i也为平稳序列。

所谓维纳滤波问题就是，令测量序列z_i通过一横向滤波器，我们期望在滤波器的输出端得到信号d的合理估计，d可以是x本身，也可以是x的线性函数，d称为期望信号。横向滤波器的权根据最小均方误差准则确定，滤波器的输出是$\hat{d}$，均方误差最小准则就是使$J = E[(d - \hat{d})^2]$达到最小，这时$\hat{d}$就称为维纳滤波。这个横向滤波器就称为维纳滤波器。

对于如(4-1-1)式所示的测量序列，期望信号d就取x_i，容易理解到，对于上述平稳随机过程的估计与第三章中的对随机变量的最小均方误差估计没有什么差别，所以，第三章中的最小均方误差估计方法可以用来解决平稳随机过程的维纳滤波问题。

回顾一下，第三章中关于最小均方误差估计的结论是：

对于测量集合$\{z_i\}$，$z_i = x_i + e_i$，$i = 1, 2, \cdots, N$，可得出最小均方误差估计为

$$\hat{x}_{lmse} = A^T_{lmse} Z \tag{4-1-2}$$

其中Z为测量矢量，$Z = [Z_1, Z_2, \cdots, Z_N]^T$，$A^T_{lmse}$为最佳权矢量，$A^T_{lmse} = [a_1, a_2, \cdots, a_N]$

$$A_{lmse} = \Phi^{-1}_{zz} \Phi_{zx} \tag{4-1-3}$$

最小均方误差为

$$P_{lmse}=E[x^2]-\Phi_{zx}^T\Phi_{zz}^{-1}\Phi_{zx} \qquad (4\text{-}1\text{-}4)$$

其中Φ_{zz}为矢量Z的自相关矩阵

$$\Phi_{zz}=E[ZZ^T] \qquad (4\text{-}1\text{-}5)$$

Φ_{zx}为Z与x的互相关矢量（当x为矢量时，Φ_{zx}成为互相关矩阵）。

$$\Phi_{zx}=E[Zx^T]=E[Zx] \qquad (4\text{-}1\text{-}6)$$

对于平稳随机过程的测量时间序列和对随机变量的测量集合之间没有什么差别，所以，对测量序列$\{z_i\}$的维纳滤波就是对测量集合的最小均方误差估计x_{lmse}。

把(4-1-2)式改写成如下形式

$$\hat{x}=A_{lmse}^TZ=[a_1,a_2,\cdots,a_N]\begin{pmatrix} z_1\\ z_2\\ \vdots\\ z_N\end{pmatrix}$$

$$=\sum_{i=1}^{N}a_iz_i \qquad (4\text{-}1\text{-}7)$$

(4-1-7)式是横向滤波器的输出表达式，故根据(4-1-7)式构造的横向滤波器就是维纳滤波器，如图4-1-1所示。

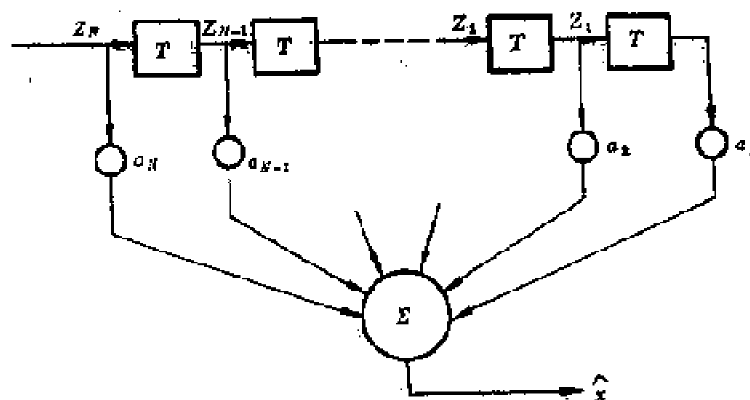


图4-1-1 维纳滤波器结构

通常横向滤波器的长度是有限的，而测量序列的长度是无限的，所以在j时刻，延迟线上的实际信号为

$$z_j,\ z_{j-1},\ \cdots,\ z_{j-N+1}$$

在j时刻，维纳滤波器的输出可表达为

$$\hat{x}_j=\sum_{i=0}^{N-1}h_iz_{j-i} \qquad (4\text{-}1\text{-}8)$$

(4-1-8)式称为卷积和，权系数h_i构成横向滤波器的冲击响应。

在j时刻维纳滤波器结构如图4-1-2所示。

若令滤波器的权为$a_1=a_2,\ \cdots,\ a_N=\dfrac{1}{N}$，则$\hat{x}=\dfrac{1}{N}\sum\limits_{i=1}^{N}z_i$，这时维纳滤波器退化为采样平均器。若令权矢量为$A=\Phi_{zz}^{-1}\Phi_{zx}$，那末它就是一个维纳滤波器，输出误差的均方值$E[(\hat{x}-x)^2]$将达到最小。

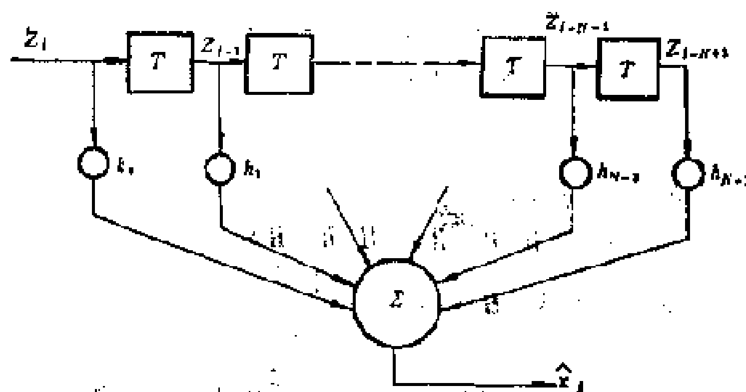


图4-1-2　在j时刻的维纳滤波器

从以上分析知道，要构造维纳滤波器，必先求出横向滤波器的最佳权矢量A_{lmse}。若已知自相关矩阵Φ_{zz}和互相关矢量Φ_{zx}中的所有元素，则立即可求出最佳权矢量为

$$A_{lmse}=\Phi_{zz}^{-1}\Phi_{zx} \tag{4-1-9}$$

矢量A_{lmse}中的元素就是横向滤波器中需要的权，于是有

$$\hat{x}_{lmse}=\Phi_{zx}^{T}\Phi_{zz}^{-1}Z. \tag{4-1-10}$$

例4-1-1　设有平稳测量序列

$$z_i=x_i+e_i \qquad i=1,2,\cdots,N \tag{4-1-11}$$

已知　$E[x_i^2]=S$，$E[e_i]=0$，$E[e_i^2]=\sigma_e^2$，$E[x_ie_j]=0$，$E[e_ie_j]=0$

试设计一个维纳滤波器，并求对x_i的估计。

解：记

$$\boldsymbol{Z}=\begin{pmatrix}z_1\\z_2\\\vdots\\z_N\end{pmatrix},\ \boldsymbol{H}=\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix},\ \boldsymbol{e}=\begin{pmatrix}e_1\\e_2\\\vdots\\e_N\end{pmatrix},\ R=E[ee^T]$$

可得矢量方程　$$\boldsymbol{Z}=\boldsymbol{H}x+\boldsymbol{e} \tag{4-1-12}$$

因为　$$\boldsymbol{A}_{lmse}^T=\Phi_{zx}^T\Phi_{zz}^{-1} \tag{4-1-13}$$

$$\Phi_{zx}=E[(\boldsymbol{H}x+\boldsymbol{e})x]=\boldsymbol{H}\boldsymbol{S} \tag{4-1-14}$$

$$\Phi_{zz}=E[(\boldsymbol{H}x+\boldsymbol{e})(\boldsymbol{H}x+\boldsymbol{e})^T]=\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}^T+\boldsymbol{R} \tag{4-1-15}$$

故有　$$\boldsymbol{A}_{lmse}^T=\boldsymbol{S}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}^T+\boldsymbol{R})^{-1} \tag{4-1-16}$$

由附录D中矩阵等式可得

$$\boldsymbol{A}_{lmse}^T=(\boldsymbol{S}^{-1}+\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{R}^{-1}$$

由于S为标量，再把$\boldsymbol{H}$、$\boldsymbol{R}$之值代入，得

$$\boldsymbol{A}_{lmse}^T=\frac{1}{N+\frac{\sigma_e^2}{S}}[1,1,\cdots,1] \tag{4-1-17}$$

所以，维纳滤波器的最佳权系数为

$$a_1=a_2=\cdots=a_N=\frac{1}{N+\frac{\sigma_e^2}{S}} \tag{4-1-18}$$

期望信号x的估计为

$$\hat{x}_{lmse} = A_{lmse}^T Z = \frac{1}{N + \frac{\sigma_e^2}{S}} \sum_{i=1}^{N} z_i \tag{4-1-19}$$

最小均方误差为

$$\begin{aligned} P_{lmse} &= S - \Phi_{zx}^T \Phi_{zz}^{-1} \Phi_{zx} \\ &= S - (S^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} H S \\ &= (S^{-1} + H^T R^{-1} H)^{-1} = \frac{\sigma_e^2}{N + \frac{\sigma_e^2}{S}} \end{aligned} \tag{4-1-20}$$

可见，本例中维纳滤波器的权系数是相同的。只需要知道输入功率信噪比s/σ_e^2就可设计相应的维纳滤波器。

有色噪声时的维纳滤波

上面分析中假设各时刻的测量噪声是不相关的，现在再来研究测量噪声相关时的情况。

今有测量序列$z_i = x_i + e_i$，其中噪声e_i为一阶马尔科夫序列

$$e_i = c e_{i-1} + v_i \tag{4-1-21}$$

其中v_i为零均值的噪声序列；c为常数，$|c| < 1$；不难求得

$$\begin{aligned} \sigma_e^2 &= \phi_e(0) = E[e_i^2] \\ &= E[(c e_{i-1} + v_i)^2] = c^2 \phi_e(0) + \phi_v(0) = c^2 \sigma_e^2 + \sigma_v^2 \end{aligned} \tag{4-1-22}$$

其中考虑了$E[e_{i-1} v_i] = 0$

故得

$$\phi_e(0) = \sigma_e^2 = \frac{\sigma_v^2}{1 - c^2}$$

$$\phi_e(1) = E[e_{i-1} e_i] = c \sigma_e^2$$

$$\phi_e(2) = E[e_{i-2} e_i] = c^2 \sigma_e^2$$

....

由于自相关函数为偶函数，故得

$$\phi_e(k) = c^{|k|} \sigma_e^2 = \frac{c^{|k|}}{1 - c^2} \sigma_v^2 \tag{4-1-23}$$

再假定已知，$E[x_i] = 0$，$E[x_i^2] = S$，$E[x_i e_j] = 0$，观测值有两个z_1、z_2，求信号x的最小均方误差估计(即维纳滤波)。

根据维纳-霍夫方程，有

$$\Phi_{zz} h_{lmse} = \Phi_{zx}$$

或

$$\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} \phi_{z_1 x} \\ \phi_{z_2 x} \end{pmatrix} \tag{4-1-24}$$

根据已知条件，容易求出

$$\left.\begin{aligned} \phi_{z_1 x} &= \phi_{z_2 x} = E[z_i x_i] = S \\ \phi_{11} &= \phi_{22} = E[z_i^2] = S + \sigma_e^2 \\ \phi_{12} &= \phi_{21} = E[z_{i+1} z_i] = S + c\sigma_e^2 \end{aligned}\right\} \tag{4-1-25}$$

于是求得

$$h_1 = h_2 = \frac{S}{2S + (1+c)\sigma_e^2} \qquad (4\text{-}1\text{-}26)$$

$$\hat{x}_{lmse} = \frac{z_1 + z_2}{2 + \frac{(1+c)\sigma_e^2}{S}} \qquad (4\text{-}1\text{-}27)$$

求估计的最小均方误差，得

$$\begin{aligned} P_{lmse} &= E[(x - \hat{x}_{lmse})^2] = E[(x - h_1 z_1 - h_2 z_2)^2] \\ &= S^2 + h_1^2 \Phi_{11} + h_2^2 \Phi_{22} + 2h_1^2 \Phi_{12} - 2h_1 \Phi_{z_1 x} - 2h_1 \Phi_{z_2 x} \\ &= 2h_1^2 (1+c)\sigma_e^2 + (2h_1 - 1)^2 S \end{aligned} \qquad (4\text{-}1\text{-}28)$$

若$c = 0$，即e_i为白噪声序列，则

$$\begin{aligned} P_{lmse} &= 2h_1^2 \sigma_e^2 + (2h_1 - 1)^2 S \\ &= \frac{\sigma_e^2}{2 + \frac{\sigma_e^2}{S}} \end{aligned} \qquad (4\text{-}1\text{-}29)$$

可见，由于噪声是有色的，使估计质量下降。

对于维纳滤波器再作几点补充说明：

1．若期望信号不是随机变量x本身而是x的某个线性函数d，则相应的维纳滤波器为

$$\boldsymbol{A}_{lmse} = \Phi_{zz}^{-1} \Phi_{zd} \qquad (4\text{-}1\text{-}30)$$

$$\hat{d}_{lmse} = \boldsymbol{A}^T_{lmse} \boldsymbol{Z} \qquad (4\text{-}1\text{-}31)$$

$$P_{lmse} = E[d^2] - \Phi_{zd}^T \Phi_{zz}^{-1} \Phi_{zd} \qquad (4\text{-}1\text{-}32)$$

如果期望信号d与测量矢量$\boldsymbol{Z}$不相关，则有$\Phi_{zd} = 0$，我们将得不到d的估计。

2．求最佳权系数时，令Φ_{zz}中的各元素为已知，事实上，Φ_{zz}为统计量，其中元素均为未知，但是，当测量z_i为各态历经的平稳随机序列时，可以用时间平均代替统计平均。求得元素$\Phi_{zz}(j)$的估计的一种方法是取

$$\hat{\hat{\Phi}}_{zz}(j) = \frac{1}{N - |j|} \sum_{i=0}^{N-|j|-1} z_i z_{i+j} \quad 0 \leqslant j \leqslant N-1 \quad N\text{为数据长度} \qquad (4\text{-}1\text{-}33)$$

两边取统计平均，显然有

$$E[\hat{\hat{\Phi}}_{zz}(j)] = \Phi_{zz}(j) \qquad (4\text{-}1\text{-}34)$$

还可以证明 $\lim\limits_{N\to\infty} \text{var}[\hat{\hat{\Phi}}_{zz}(j)] \to 0 \qquad j \ll N$

所以，$\hat{\hat{\Phi}}_{zz}(j)$是$\Phi_{zz}(j)$的无偏一致估计量。

另一种方法是取

$$\hat{\Phi}_{zz}(j) = \frac{1}{N} \sum_{i=0}^{N-|j|-1} z_i z_{i+j} \qquad (4\text{-}1\text{-}35)$$

由于 $\hat{\Phi}_{zz}(j) = \frac{N-|j|}{N} \hat{\hat{\Phi}}_{zz}(j)$

所以，$\hat{\Phi}_{zz}(j)$是$\Phi_{zz}(j)$的有偏估计，但它是渐近无偏的，且偏量

$$b = \left| \frac{j}{N} \right| \Phi_{zz}(j) \qquad (4\text{-}1\text{-}36)$$

注意到当j接近记录长度N时，只有少数几个样点可用来计算$\hat{\hat{\Phi}}_{zz}(j)$，$\mathrm{var}[\hat{\hat{\Phi}}_{zz}(j)]$将变得非常大，估计质量下降。与此相反，有偏估计$\hat{\Phi}_{zz}(j)$的方差不会变得如此之大，但是当$j$接近$N$时，偏量$b$接近于$\Phi_{zz}(j)$本身。

以上是把记录长度N固定，改变j来比较两种估计的性能的，再令j固定，改变N来比较两种估计。

由于这两种估计都是$\Phi_{zz}(j)$的一致估计量，但$\hat{\Phi}_{zz}(j)$的方差要小得多，且偏量b也随N的增大而下降，可以预期，$\hat{\Phi}_{zz}(j)$的均方误差要小于$\hat{\hat{\Phi}}_{zz}(j)$的均方误差，所以一般情况下，取(4-1-35)式作为计算$\Phi_{zz}(j)$的基础，并且总希望增加记录长度N来改善相关函数的估计。

3. 若期望信号d为N维矢量，此时Φ_{zd}成为互相关矩阵，维纳滤波器的权将成为$N\times N$维矩阵

$$A_{lmse}=\Phi_{zz}^{-1}\Phi_{zd}$$

4. 当缺乏期望信号d的先验知识时，不能求得互相关矢量Φ_{zd}，使得维纳滤波器的设计发生了困难。

5. 当期望信号是一平稳机序列时，互相关Φ_{zd}为一矢量，相应的权A_{lmse}也是一矢量，故维纳滤波器的结构为横向滤波器。

当期望信号为非平稳时，互相关函数为

$$\Phi_{zd}=E[\boldsymbol{ZD}]$$

$$=E\left\{\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix}[d_1,\ d_2,\ \cdots,d_N]\right\}$$

这时Φ_{zd}为矩阵，故相应的权A_{lmse}也是一矩阵。一般情况下，不能求得矩阵Φ_{zd}和Φ_{zz}中各元素的估计，因此说，维纳滤波只适用于平稳过程的估计，而不适用于非平稳过程的估计。

§4-2 维纳滤波器权系数的递推算法

若利用N个测量数据，已经算出了维纳滤波器的权系数矢量A^N，$A^N=[a_1^N,a_2^N,\cdots,a_N^N]^T$，在新增加一个测量值$z_{N+1}$之后，需要重新计算权矢量$A^{N+1}$，$A^{N+1}=[a_1^{N+1},a_2^{N+1},\cdots,a_{N+1}^{N+1}]^T$，这是一件很费时间的事。如果能采用递推算法，由$A^N$递推算出$A^{N+1}$，那么就可以缩短计算时间。

在推导权矢量A^N的递推算法之前，先来推导一种特殊形式的陶布利兹方程的递推解法。

设有陶布里兹方程如下：

$$\begin{bmatrix} r_1 & r_2\cdots r_n \\ r_2 & r_1\cdots r_{n-1} \\ \cdots & \cdots \\ r_n & r_{n-1}\cdots r_1 \end{bmatrix}\begin{bmatrix} a_1^{(n)} \\ a_2^{(n)} \\ \vdots \\ a_n^{(n)} \end{bmatrix}=\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{4-2-1}$$

上式左边的矩阵称为陶布里兹矩阵，其中只有n个独立元素，假若已经求出方程式的解为

$$a_1^{(n)},\ a_2^{(n)},\ \cdots,\ a_n^{(n)}$$

今有$n+1$阶陶布里兹方程为

$$\begin{pmatrix} r_1 & r_2 & \cdots r_n & r_{n+1} \\ r_2 & r_1 & \cdots r_{n-1} & r_n \\ \cdots & \cdots & \cdots & \cdots \\ r_n & r_{n-1} & \cdots r_1 & r_2 \\ r_{n+1} & r_n & \cdots r_2 & r_1 \end{pmatrix} \begin{pmatrix} a_1^{(n+1)} \\ a_2^{(n+1)} \\ \vdots \\ a_n^{(n+1)} \\ a_{n+1}^{(n+1)} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \tag{4-2-2}$$

其中r_{n+1}为已知，要求根据已知的n阶方程的解$a_1^{(n)}, a_2^{(n)}, \cdots, a_n^{(n)}$，递推地计算出（4-2-2)式的解$a_1^{(n+1)}$，$a_2^{(n+1)}$，$\cdots$，$a_{n+1}^{(n+1)}$。

首先把(4-2-1)式扩充为$n+1$阶，得

$$\begin{pmatrix} r_1 & r_2 & \cdots r_n & r_{n+1} \\ r_2 & r_1 & \cdots r_{n-1} & r_n \\ \cdots & \cdots & \cdots & \cdots \\ r_n & r_{n-1} & \cdots r_1 & r_2 \\ r_{n+1} & r_n & \cdots r_2 & r_1 \end{pmatrix} \begin{pmatrix} a_1^{(n)} \\ a_2^{(n)} \\ \vdots \\ a_n^{(n)} \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ \rho_n \end{pmatrix} \tag{4-2-3}$$

其中

$$\rho_n = r_{n+1} a_1^{(n)} + r_n a_2^{(n)} + \cdots + r_2 a_n^{(n)} \tag{4-2-4}$$

(4-2-4)式与(4-2-3)式是等价的，把(4-2-3)改写成另一种等价形式

$$\begin{pmatrix} r_1 & r_2 & \cdots r_n & r_{n+1} \\ r_2 & r_1 & \cdots r_{n-1} & r_n \\ \cdots & \cdots & \cdots & \cdots \\ r_n & r_{n-1} & \cdots r_1 & r_2 \\ r_{n-1} & r_n & \cdots r_2 & r_1 \end{pmatrix} \begin{pmatrix} 0 \\ a_n^{(n)} \\ \vdots \\ a_2^{(n)} \\ a_1^{(n)} \end{pmatrix} = \begin{pmatrix} \rho_n \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \tag{4-2-5}$$

用ρ_n乘(4-2-5)式两边，再从(4-2-3)式减去ρ_n乘(4-2-5)式，得

$$\begin{pmatrix} r_1 & r_2 & \cdots r_n & r_{n+1} \\ r_2 & r_1 & \cdots r_{n-1} & r_n \\ \cdots & \cdots & \cdots & \cdots \\ r_n & r_{n-1} & \cdots r_1 & r_2 \\ r_{n+1} & r_n & \cdots r_2 & r_1 \end{pmatrix} \begin{pmatrix} a_1^{(n)} \\ a_2^{(n)} - \rho_n a_n^{(n)} \\ \vdots \\ a_n^{(n)} - \rho_n a_2^{(n)} \\ -\rho_n a_1^{(n)} \end{pmatrix} = \begin{pmatrix} 1-\rho_n^2 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \tag{4-2-6}$$

因为$1-\rho_n^2 \neq 0$。用$\frac{1}{1-\rho_n^2}$乘(4-2-6)式两边，然后再与(4-2-2)式比较，便得到

$$\begin{cases} a_1^{(n+1)} = \dfrac{1}{1-\rho_n^2} a_1^{(n)} \\ a_2^{(n+1)} = \dfrac{1}{1-\rho_n^2} \left(a_2^{(n)} - \rho_n a_n^{(n)}\right) \\ \cdots\cdots\cdots\cdots \\ a_n^{(n-1)} = \dfrac{1}{1-\rho_n^2} \left(a_n^{(n)} - \rho_n a_2^{(n)}\right) \\ a_{n+1}^{(n+1)} = \dfrac{1}{1-\rho_n^2} \left(-\rho_n a_1^{(n)}\right) \end{cases} \tag{4-2-7}$$

当$n=1$时，由(4-2-1)得$r_1 a_1^{(1)} = 1$，即有

$$a_1^{(1)} = \frac{1}{r_1}$$

有了初值$a_1^{(1)}$便可根据(4-2-4)式与(4-2-7)式求得$n=2$时的解$a_1^{(2)}$和$a_2^{(2)}$，依此类推，可求得任意n值时的解，$a_1^{(n)}$，$a_2^{(n)}$，…，$a_n^{(n)}$。

现在，回过头来再求维纳-霍夫方程的递推解法，首先写出n阶维纳-霍夫方程内

$$\Phi_{zz}^{(n)}h^{(n)}=\Phi_{zd}^{(n)}$$

写成展开形式

记
$$h^{(n)}=\begin{pmatrix}h_1^{(n)}\\h_2^{(n)}\\\vdots\\h_n^{(n)}\end{pmatrix},\quad \Phi_{zd}^{(n)}=\begin{pmatrix}g_1\\g_2\\\vdots\\g_n\end{pmatrix}$$

则有
$$\begin{pmatrix}\Phi_0 & \Phi_1 & \cdots\Phi_{n-1}\\ \Phi_1 & \Phi_0 & \cdots\Phi_{n-2}\\ \cdots & \cdots & \cdots\\ \Phi_{n-1} & \Phi_{n-2} & \cdots\ \Phi_0\end{pmatrix}\begin{pmatrix}h_1^{(n)}\\h_2^{(n)}\\\vdots\\h_n^{(n)}\end{pmatrix}=\begin{pmatrix}g_1\\g_2\\\vdots\\g_n\end{pmatrix}\tag{4-2-8}$$

再写出$n+1$阶维纳-霍夫方程

$$\begin{pmatrix}\Phi_0 & \Phi_1 & \cdots\Phi_{n-1} & \Phi_n\\ \Phi_1 & \Phi_0 & \cdots\Phi_{n-2} & \Phi_{n-1}\\ \cdots & \cdots & \cdots & \cdots\\ \Phi_{n-1} & \Phi_{n-2}\cdots\Phi_0 & & \Phi_1\\ \Phi_n & \Phi_{n-1}\cdots\Phi_1 & & \Phi_0\end{pmatrix}\begin{pmatrix}h_1^{(n+1)}\\h_2^{(n+1)}\\\vdots\\h_n^{(n+1)}\\h_{n+1}^{(n+1)}\end{pmatrix}=\begin{pmatrix}g_1\\g_2\\\vdots\\g_n\\g_{n+1}\end{pmatrix}\tag{4-2-9}$$

其中　$\Phi_n=\Phi_{j-i}=E[z_iz_j]$

$i,j=1,2,\cdots,n+1$

现在希望根据已知n阶维纳-霍夫方程的解，$h_1^{(n)},h_2^{(n)}$，…，$h_n^{(n)}$，在得到Φ_n之后，再求出$n+1$阶维纳-霍夫方程的解$h_1^{(n+1)}$，$h_2^{(n+1)}$，…，$h_{n+1}^{(n+1)}$。

把(4-2-8)式扩充为$n+1$阶，得到其等价形式为

$$\begin{pmatrix}\Phi_0 & \Phi_1 & \cdots\Phi_{n-1} & \Phi_n\\ \Phi_1 & \Phi_0 & \cdots\Phi_{n-2} & \Phi_{n-1}\\ \cdots & \cdots & \cdots & \cdots\\ \Phi_{n-1} & \Phi_{n-2}\cdots\Phi_0 & & \Phi_1\\ \Phi_n & \Phi_{n-1}\cdots\Phi_1 & & \Phi_0\end{pmatrix}\begin{pmatrix}h_1^{(n)}\\h_2^{(n)}\\\vdots\\h_n^{(n)}\\0\end{pmatrix}=\begin{pmatrix}g_1\\g_2\\\vdots\\g_n\\\beta_n\end{pmatrix}\tag{4-2-10}$$

其中
$$\beta_n=\Phi_nh_1^{(n)}+\Phi_{n-1}h_2^{(n)}+\cdots+\Phi_1h_n^{(n)}\tag{4-2-11}$$

从(4-2-10)式减去(4-2-9)式，得

$$\begin{pmatrix}\Phi_0 & \Phi_1 & \cdots\Phi_n & \Phi_{n-1}\\ \Phi_1 & \Phi_0 & \cdots\Phi_{n-2} & \Phi_{n-1}\\ \cdots & \cdots & \cdots & \cdots\\ \Phi_{n-1} & \Phi_{n-2}\cdots\Phi_0 & & \Phi_1\\ \Phi_n & \Phi_{n-1}\cdots\Phi_1 & & \Phi_0\end{pmatrix}\begin{pmatrix}h_1^{(n+1)}-h_1^{(n)}\\h_2^{(n+1)}-h_2^{(n)}\\\vdots\\h_n^{(n+1)}-h_n^{(n)}\\h_{n+1}^{(n+1)}\end{pmatrix}=\begin{pmatrix}0\\0\\\vdots\\0\\g_{n+1}-\beta_n\end{pmatrix}\tag{4-2-12}$$

(4-2-12)式两边除以$(g_{n+1}-\beta_n)$，再与(4-2-2)式比较，便得到

$$\left.\begin{array}{l}h_1^{(n+1)}=h_1^{(n)}+(g_{n+1}-\beta_n)a_{n+1}^{(n+1)}\\ h_2^{(n+1)}=h_2^{(n)}+(g_{n+1}-\beta_n)a_n^{(n+1)}\\ \qquad\vdots\\ h_n^{(n+1)}=h_n^{(n)}+(g_{n+1}-\beta_n)a_2^{(n+1)}\\ \qquad h_{n+1}^{(n+1)}=(g_{n+1}-\beta_n)a_1^{(n+1)}\end{array}\right\} \tag{4-2-13}$$

递推从$n=1$开始，由(4-2-1)式和(4-2-8)式得

$$a_1^{(1)}=\frac{1}{\Phi_0}\qquad h_1^{(1)}=\frac{g_1}{\Phi_0} \tag{4-2-14}$$

根据已知的$\Phi_0,\Phi_1,\Phi_2,\cdots,\Phi_n$和$g_1$，$g_2$，$\cdots$，$g_n$，$g_{n+1}$就可以进行递推计算，仿照(4-2-2)式的解法，就可得出$n+1$阶维纳-霍夫方程的解为

$$\left.\begin{array}{l}\rho_n=\phi_n a_1^{(n)}+\phi_{n-1}a_2^{(n)}+\cdots+\phi_1 a_n^{(n)}\\ \beta_n=\phi_n h_1^{(n)}+\phi_{n-1}h_2^{(n)}+\cdots+\phi_1 h_n^{(n)}\\ a_1^{(n+1)}=\dfrac{1}{1-\rho_n^2}a_1^{(n)}\\ a_2^{(n+1)}=\dfrac{1}{1-\rho_n^2}(a_2^{(n)}-\rho_n a_n^{(n)})\\ \cdots\cdots\cdots\cdots\\ a_n^{(n+1)}=\dfrac{1}{1-\rho_n^2}(a_n^{(n)}-\rho_n a_n^{(n)})\\ a_{n+1}^{(n+1)}=\dfrac{1}{1-\rho_n^2}(-\rho_n a_1^{(n)})\\ h_1^{(n+1)}=h_1^{(n)}+(g_{n+1}-\beta_n)a_{n+1}^{(n+1)}\\ h_2^{(n+1)}=h_2^{(n)}+(g_{n+1}-\beta_n)a_n^{(n+1)}\\ \cdots\cdots\cdots\cdots\\ h_n^{(n+1)}=h_n^{(n)}+(g_{n+1}-\beta_n)a_2^{(n+1)}\\ h_{n+1}^{(n+1)}=(g_{n+1}-\beta_n)a_1^{(n+1)}\end{array}\right\} \tag{4-2-15}$$

如果已知相关函数ϕ_0，ϕ_1，ϕ_2，$\cdots$，ϕ_n之值，就可从$n=1$开始，递推地求出$n+1$阶维纳-霍夫方程的解$h_1^{(n+1)}$，$h_2^{(n+1)}$，$\cdots$，$h_n^{(n+1)}$，$h_{n+1}^{(n+1)}$。

§4-3　莱文森预测滤波算法

假使已得到平稳随机过程的离散测量序列z_0，z_1，z_2，$\cdots$，z_{n-1}。希望对z_n作出预报估计，此预报估计以符号$\hat{z}_{n/n-1}$表示之。

为了得到此预报估计，仍采用最小均方误差估计方法，令$\hat{z}_{n/n-1}$为以前测量值的加权和

$$\hat{z}_{n/n-1}=\sum_{i=0}^{n-1}a_{(n-i)}^{(n)}z_i \tag{4-3-1}$$

其中$a_{n-i}^{(n)}$为n阶滤波器的权系数，根据正交原理，得出维纳-霍夫方程为

$$E[(z_n-\hat{z}_{n/n-1})z_j]=0\qquad j=0,1,\cdots,n-1 \tag{4-3-2}$$

或 $$E[z_n z_j]=\sum_{i=0}^{n-1}a_{n-i}^{(n)}E[z_i z_j] \qquad j=0,1,\cdots,n-1 \tag{4-3-3}$$

对于平稳机过程而言，有

$$E[z_i z_j]=\Phi_{j-i}$$
$$E[z_j z_0]=\Phi_j$$
$$\Phi_j=\Phi_{-j}$$

因此，(4-3-3)式可写成

$$\Phi_{n-j}=\sum_{i=0}^{n-1}a_{n-i}^{(n)}\Phi_{j-i} \qquad j=0,1,\cdots,n-1 \tag{4-3-4}$$

若已知相关系数 Φ_{j-i} 之值，则由(4-3-4)式不难解出预测滤波器的各权系数，再由(4-3-1)式算出预报估计 $\hat{z}_{n/n-1}$。因为(4-3-4)式为一联立方程组，计算起来比较麻烦，特别当数据长度逐步增长时，重复计算更加费力费时，莱文森提出了一种递推算法，可使计算量大大下降，下面就来研究莱文森预测算法。

令预测估计的均方误差为 J_n

$$J_n=E[(z_n-\hat{z}_{n/n-1})^2]$$
$$=E[(z_n-\hat{z}_{n/n-1})]z_n-E[(z_n-\hat{z}_{n/n-1})\hat{z}_{n/n-1}] \tag{4-3-5}$$

根据正交原理，得知估计误差与估计正交，即有

$$E[(z_n-\hat{z}_{n/n-1})\hat{z}_{n/n-1}]=0 \tag{4-3-6}$$

故 $$J_n=E[(z_n-\hat{z}_{n/n-1})z_n]$$
$$=E[z_n^2]-\sum_{i=0}^{n-1}a_{n-i}^{(n)}E[z_i z_n]$$
$$=\Phi_0-\sum_{i=0}^{n-1}a_{n-i}^{(n)}\Phi_{n-i} \tag{4-3-7}$$

可见均方误差 J_n 是随 n 的上升而单调下降的，J_n 的值可按 $n=1,2,\cdots$ 顺序计算，把计算到的 J_n 与一预定门限相比较，一直到接近门限为止。所以，希望有一个递推计算 a_{n-i}^{n} 和 J_n 的算法，以便设计出一个均方误差不大于门限的预测滤波器，莱文森预测滤波器的结构如图4-3-1所示，形式上与维纳滤波器相似，但权系数不同。

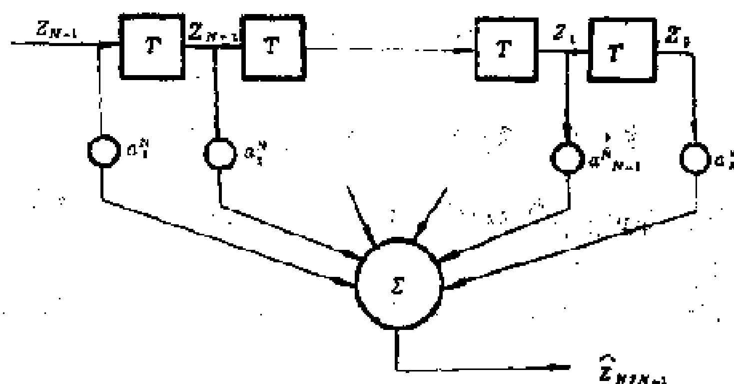


图4-3-1 莱文森预测滤波器结构

考虑到在标量情况下有

$$E[z_i z_j]=E[z_j z_i]$$

把(4-3-4)式与(4-3-7)式合并，得

$$[1, -a_1^{(n)}, -a_2^{(n)}, \cdots, -a_n^{(n)}]\begin{pmatrix} \Phi_0 & \Phi_1 & \cdots\Phi_n \\ \Phi_1 & \Phi_0 & \cdots\Phi_{n-1} \\ \cdots & \cdots & \cdots \\ \Phi_n & \Phi_{n-1} & \cdots\Phi_0 \end{pmatrix} = [J_n, 0, \cdots, 0] \tag{4-3-8}$$

对(4-3-8)式加以转置，得

$$\begin{pmatrix} \Phi_0 & \Phi_1 & \cdots\Phi_n \\ \Phi_1 & \Phi_0 & \cdots\Phi_{n-1} \\ \cdots & \cdots & \cdots \\ \Phi_n & \Phi_{n-1} & \cdots\Phi_0 \end{pmatrix}\begin{pmatrix} 1 \\ -a_1^{(n)} \\ \vdots \\ -a_n^{(n)} \end{pmatrix} = \begin{pmatrix} J_n \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{4-3-9}$$

假设$a_1^{(n)}, a_2^{(n)}, \cdots, a_n^{(n)}$和$J_n$之值已经求出,我们要寻求$a_1^{(n+1)}, a_2^{(n+1)}, \cdots, a_{n+1}^{(n+1)}$和$J_{n+1}$。

注意到这正是(4-2-1)式类型的陶布里兹方程，对于$n+2$阶陶布里兹方程，我们有

$$\begin{pmatrix} \Phi_0 & \Phi_1 & \cdots\Phi_n & \Phi_{n+1} \\ \Phi_1 & \Phi_0 & \cdots\Phi_{n-1} & \Phi_n \\ \cdots & \cdots & \cdots & \cdots \\ \Phi_n & \Phi_n & \cdots\Phi_0 & \Phi_1 \\ \Phi_{n+1} & \Phi_n & \cdots\Phi_1 & \Phi_0 \end{pmatrix}\begin{pmatrix} 1 \\ -a_1^{(n+1)} \\ \vdots \\ -a_n^{(n+1)} \\ -a_{n+1}^{(n+1)} \end{pmatrix} = \begin{pmatrix} J_{n+1} \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \tag{4-3-10}$$

其中$a_i^{(n)}$和J_n为已知，$i=0,1,\cdots,n-1$

记
$$\rho_n = \Phi_{n+1} - \sum_{i=1}^{n} a_i^{(n)}\Phi_{n-i+1} \tag{4-3-11}$$

把(4-3-9)式和(4-3-11)式合并，得

$$\begin{pmatrix} \Phi_0 & \Phi_1 & \cdots\cdots\Phi_{n+1} \\ \Phi_1 & \Phi_0 & \cdots\cdots\Phi_n \\ \cdots & \cdots & \cdots \\ \Phi_n & \Phi_{n-1} & \cdots\cdots\Phi_1 \\ \Phi_{n+1} & \Phi_n & \cdots\cdots\Phi_0 \end{pmatrix}\begin{pmatrix} 1 \\ -a_1^{(n)} \\ \vdots \\ -a_n^{(n)} \\ 0 \end{pmatrix} = \begin{pmatrix} J_n \\ 0 \\ \vdots \\ \\ \rho_n \end{pmatrix} \tag{4-3-12}$$

根据陶布里兹方程的性质，把(4-3-12)式改写成

$$\begin{pmatrix} \Phi_0 & \Phi_1 & \cdots\Phi_{n+1} \\ \Phi_1 & \Phi_0 & \cdots\Phi_n \\ \cdots & \cdots & \cdots \\ \Phi_n & \Phi_{n-1} & \cdots\Phi_1 \\ \Phi_{n-1} & \Phi_{n-2} & \cdots\Phi_0 \end{pmatrix}\begin{pmatrix} 0 \\ -a_n^{(n)} \\ \vdots \\ -a_1^{(n)} \\ 1 \end{pmatrix} = \begin{pmatrix} \rho_n \\ 0 \\ \vdots \\ 0 \\ J_n \end{pmatrix} \tag{4-3-13}$$

对(4-3-13)式两边乘以$\left(-\frac{\rho_n}{J_n}\right)$，再与(4-3-12)式相加，得

$$
\begin{pmatrix}
\Phi_0 & \Phi_1 & \cdots \Phi_{n+1} \\
\Phi_1 & \Phi_2 & \cdots \Phi_n \\
\cdots & \cdots & \cdots \\
\Phi_n & \Phi_{n-1} & \cdots \Phi_1 \\
\Phi_{n+1} & \Phi_n & \cdots \Phi_0
\end{pmatrix}
\begin{pmatrix}
1 \\
-a_1^{(n)} + \dfrac{\rho_n}{J_n} a_n^{(n)} \\
\vdots \\
-a_n^{(n)} + \dfrac{\rho_n}{J_n} a_1^{(n)} \\
-\dfrac{\rho_n}{J_n}
\end{pmatrix}
=
\begin{pmatrix}
J_n - \dfrac{\rho_n^2}{J_n} \\
0 \\
\vdots \\
0 \\
0
\end{pmatrix}
\tag{4-3-14}
$$

对比(4-3-13)式与(4-3-14)式，得

$$
\left.
\begin{aligned}
&a_i^{(n+1)} = a_i^{(n)} - \frac{\rho_n}{J_n} a_{n+1-i}^{(n)} \\
&a_{n+1}^{(n+1)} = \frac{\rho_n}{J_n} \\
&J_{n+1} = J_n - \frac{\rho_n^2}{J_n}
\end{aligned}
\right\}
\tag{4-3-15}
$$

其中
$$
\rho_n = \Phi_{n+1} - \sum_{i=1}^{n} a_i^{(n)} \Phi_{n-i+1}
$$

从$n=0$开始，初始值由(4-3-7)式得出

$$
J_1 = \Phi_0 - \frac{\Phi_1^2}{\Phi_0} \qquad a_{1(1)} = \frac{\Phi_1}{\Phi_0} \tag{4-3-16}
$$

因为J_n是随n单调下降的，根据已知的Φ_0，Φ_1，Φ_2，…逐步增大n，直到算出的J_n到达某一门限为止，由最后一步得的$a_i^{(n)}$，$i=0,1,\cdots,n-1$，就是需要莱文森预测滤波器的权系数，这时得到的预测值，其均方误差不大于设定的门限。

§4-4 非平稳随机序列维纳滤波

设测量序列 $z_i=h_i x_i+e_i$，$i=1,2,\cdots,N$ 为非平稳随机序列。造成非平稳的原因有两种：第一种原因是信号x_i本身是平稳的，但h_i是时变的，噪声e_i是平稳的，故z_i为非平稳的；第二种原因是x_i本身就是非平稳的，故z_i亦必定是非平稳的。

对于第一种情况，称为时变测量，由于h_i为确知量，故可以用维纳滤波器得出期望信号x的估计；对于第二种情况，除了少数特殊情况外，一般说来，维纳滤波器不再适用，需要应用以后讲到的卡尔曼滤波器。

时变测量时的维纳滤波

设测量序列为

$$
z_i = h_i x_i + e_i \qquad i=1,2,\cdots,N \tag{4-4-1}
$$

因为信号x_i是平稳的，期望信号就是随机变量x_i，记

$$Z=\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{pmatrix},\qquad H=\begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{pmatrix},\qquad e=\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{pmatrix}$$

故得出矢量测量方程为

$$Z=Hx+e \tag{4-4-2}$$

已知 $E[e]=0,\ E[ee^T]=R$

$E[x_i^2]=S,\ E[xe_i]=0$

根据例4-1-1的结果，容易求得

维纳滤波器的权矢量为

$$A^T_{lms}=[S^{-1}+H^TR^{-1}H]^{-1}H^TR^{-1} \tag{4-4-3}$$

测量信号x的最小均方误差估计为

$$\hat{x}_{lms}=[S^{-1}+H^TR^{-1}H]^{-1}H^TR^{-1}Z \tag{4-4-4}$$

最小均方误差为

$$P_{lms}=(S^{-1}+H^TR^{-1}H)^{-1} \tag{4-4-5}$$

若测量噪声e_i是非平稳的白噪声，则

$$R=\begin{pmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots\ \sigma_N^2 \end{pmatrix} \tag{4-4-6}$$

故容易求得

$$A^T_{lms}=\frac{1}{\frac{1}{S}+\sum\limits_{i=1}^{N}\frac{h_i^2}{\sigma_i^2}}\left(\frac{h_1}{\sigma_1^2},\ \frac{h_2}{\sigma_2^2},\ \cdots,\ \frac{h_N}{\sigma_N^2}\right) \tag{4-4-7}$$

$$\hat{x}_{lms}=\frac{1}{\frac{1}{S}+\sum\limits_{i=1}^{N}\frac{h_i^2}{\sigma_i^2}}\sum_{i=1}^{N}\frac{h_i}{\sigma_i^2}z_i \tag{4-4-8}$$

$$P_{lms}=\left(\frac{1}{S}+\sum_{i=1}^{N}\frac{h_i^2}{\sigma_i^2}\right)^{-1} \tag{4-4-9}$$

例4-4-1　设有信号$z(t)=x\cos\omega t+e(t)$，其中x为平稳随机振幅，已知$E[x]=0$，$E[x^2]=\sigma_x^2$，$e(t)$为平稳零均值白噪声，方差等于σ_e^2，在$\omega t=0$，$\omega t=\frac{\pi}{4}$时刻采样，得采样值z_1、z_2。求维纳滤波器权矢量A_{lms}，估计$\hat{x}_{lms}$及其方差$P_{\hat{x}}$。

解：此题是一时变测量情况，我们有

$$S=\sigma_x^2,\qquad R=\begin{pmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_e^2 \end{pmatrix}\qquad H=\begin{pmatrix} 1 \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

代入(4-4-3)式，得

$$A^T_{lms}=[S^{-1}+H^TR^{-1}H]^{-1}H^TR^{-1}$$

$$=\left(\frac{1}{\frac{1}{\sigma_x^2}+\frac{3}{2\sigma_e^2}}\right)\frac{1}{\sigma_e^2}\left[1\quad\frac{1}{\sqrt{2}}\right]$$

再由(4-4-4)式得

$$\hat{x}_{lms}=\frac{1}{\frac{\sigma_e^2}{\sigma_x^2}+\frac{3}{2}}z_1+\frac{1}{\sqrt{2}}\frac{1}{\frac{\sigma_e^2}{\sigma_x^2}+\frac{3}{2}}z_2$$

再由(4-4-5)式得

$$P_{\hat{x}}=\frac{1}{\frac{1}{\sigma_x^2}+\frac{3}{2\sigma_e^2}}$$

期望信号时变时的维纳滤波

设测量序列为 $z_k=x_k+e_k$，$k=1,2,3,\cdots$其中 x_k 为非平稳随机序列，故在这种情况下，期望信号随时间 k 而改变，以 d_k 表示之。可以理解到，这时维纳滤波器将是一时变滤波器，这是合乎逻辑的推理。现把滤波器的权系数用符号 $h(k,i)$ 表示，它表示在 k 时刻滤波器的第 i 个权的值，于是得出非平稳随机过程的维纳–霍夫方程为

$$\sum_{i=1}^{m}h(k,i)E[z(k,i)z(k,j)]=E[x(k,i)E(k,j)] \qquad (4\text{-}4\text{-}10)$$

$$j=1,\ 2,\cdots,\ m$$

这里 m 为横向滤波器的长度，令 k 保持不变，对 $i,j=1,2,\cdots,m$ 展开，得到一组联立方程，由此可解出 k 时刻的滤波器权系数 $h(k,i)$。

非平稳过程维纳滤波器的结构示于图4-4-1。

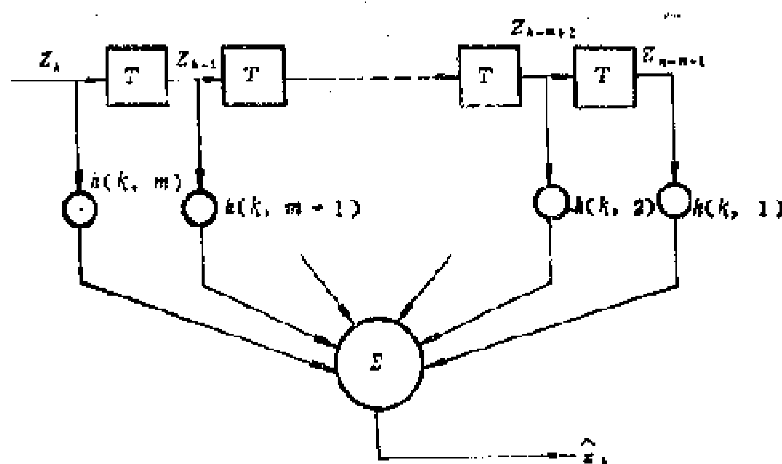


图4-4-1　非平稳过程维纳滤波器结构

在 k 时刻，测量矢量为

$$Z_k^T=[z_{k-m+1},\ z_{k-m+2},\ \cdots,\ z_k] \qquad (4\text{-}4\text{-}11)$$

滤波器的权矢量为

$$H_k^T=[h(k,1),\ h(k,2),\ \cdots,h(k,m)] \qquad (4\text{-}4\text{-}12)$$

根据正交原理，在 k 时刻维纳滤波器的权矢量应当是

$$H_k = \Phi_{zz}^{-1}(k)\Phi_{zx}(k) \tag{4-4-13}$$

从以上讨论可以看到，对非平稳时间序列进行维纳滤波，通常是十分困难的，因为

1．权系数是时变的，这是一个时变滤波器。因此，难以用硬件实现，即使用软件实现，也难以做到实时处理。

2．在每一时刻 k，要求计算自相关 $\Phi_{zz}(k)$ 和互相关 $\Phi_{zd}(k)$，这要求在每一时刻保留 m 个测量数据，并要求知道每一时刻的相关函数知识，这是难以做到的。

3．当滤波器长度改变时，必须全部重新计算一遍。

由以上所述可见，维纳滤波通常只适合于处理输入序列为平稳和期望信号为平稳的情况，这时，维纳滤波器是时不变的。否则，维纳滤波器将是时变系统。因此，维纳滤波器的应用受到极大限制。

但是对于某些特殊情况，非平稳过程的维纳滤波器也是可以得到的，下面给出用维纳滤波器求落体的高度和速度的例子。

维纳滤波应用于落体

一个落体的高度和速度都是时变的，测量只是落体高度，要求对落体的高度和速度同时作出估计。显然，这是一个时变矢量信号的估计问题。

设高度测量数据为 $z(1)$，$z(2)$，$\cdots$，$z(m)$，记 $x_1(k)$ 表示落体高度，$x_2(k)$ 表示落体速度，则期望矢量信号为

$$X(k)=\begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix}=\begin{pmatrix} x(k) \\ \dot{x}(k) \end{pmatrix} \tag{4-4-14}$$

相应的维纳-霍夫方程为

$$\sum_{i=1}^{m} H(k,i)\cdot E[z(i)z(j)] = E[X(k)z(j)] \tag{4-4-15}$$

其中

$$E[X(k)z(j)]=\begin{pmatrix} E[x_1(k)z(j)] \\ E[x_2(k)z(j)] \end{pmatrix} =\begin{pmatrix} \Phi_{x_1z}(k,j) \\ \Phi_{x_2z}(k,j) \end{pmatrix} \tag{4-4-16}$$

记

$$\Phi_{zz}(i,j) = E[z(i)z(j)]$$

$$H(k,i)=\begin{pmatrix} H_1(k,i) \\ H_2(k,i) \end{pmatrix}$$

代入（4-4-15）式，得

$$\sum_{i=1}^{m}\begin{pmatrix} H_1(k,i) \\ H_2(k,i) \end{pmatrix}\Phi_{zz}(i,j)=\begin{pmatrix} \Phi_{x_1z}(k,j) \\ \Phi_{x_2z}(k,j) \end{pmatrix} \tag{4-4-16}$$

把上式分写成两个方程式

$$\sum_{i=1}^{m} H_1(k,i)\Phi_{zz}(i,j) = \Phi_{x_1z}(k,j) \tag{4-4-17}$$

$$\sum_{i=1}^{m} H_2(k,i)\Phi_{zz}(i,j) = \Phi_{x_2z}(k,j) \tag{4-4-18}$$

相应的估计量$\hat{x}_1(k)$、$\hat{x}_2(k)$为

$$\hat{x}_1(k) = \sum_{i=1}^{m} H_1(k,i)z(i) \tag{4-4-19}$$

$$\hat{x}_2(k) = \sum_{i=1}^{m} H_2(k,i)z(i) \tag{4-4-20}$$

估计的均方误差为

$$P_{11}(k) = E[x_1^2(k)] - \sum_{i=1}^{m} H_1(k,i)\Phi_{zx_1}(i,k) \tag{4-4-21}$$

$$P_{22}(k) = E[x_2^2(k)] - \sum_{i=1}^{m} H_2(k,i)\Phi_{zx_2}(i,k) \tag{4-4-22}$$

所以，要得出维纳-霍夫方程的解就需要知道(4-4-17)～(4-4-22)式中各相关函数。因为测量方程为

$$z(j) = x_1(j) + e(j) \tag{4-4-23}$$

故

$$\Phi_{x_1z}(k,j) = E[x_1(k)z(j)]$$
$$= E[x_1(k)x_1(j)] = \Phi_{x_1}(k,j) \tag{4-4-24}$$

$$\Phi_{x_2z}(k,j) = E[x_2(k)z(j)]$$
$$= E[x_2(k)x_1(j)] = \Phi_{x_2x_1}(k,j) \tag{4-4-25}$$

$$\Phi_{zz}(i,j) = E\{[x_1(i)+e(i)][x_1(j)+e(j)]\}$$
$$= \Phi_{x_1}(i,j) + \sigma_e^2\delta_{ij} \tag{4-4-26}$$

其中δ_{ij}为克罗内克符号

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

由于$x_1(k)$、$x_2(k)$都是标量，故有

$$\Phi_{x_1}(j,k) = \Phi_{x_1}(k,j) \tag{4-4-27}$$

$$\Phi_{x_1x_2}(j,k) = \Phi_{x_2x_1}(k,j) \tag{4-4-28}$$

显见，所有需要的各相关函数都可以用$\Phi_{x_1}(k,j)$和$\Phi_{x_2x_1}(k,j)$来表示，因此，要求解维纳-霍夫方程必先求出$\Phi_{x_1}(k,j)$和$\Phi_{x_2x_1}(k,j)$。

因为

$$\Phi_{x_1}(k,j) = E[x_1(k)x_2(j)] \tag{4-4-29}$$

$$\Phi_{x_1x_2}(k,j) = E[x_1(k)x_2(j)] \tag{4-4-30}$$

根据牛顿力学，对落体有

$$x_1(k) = x_1(0) + kTx_2(0) - \frac{(kT)^2}{2}g \tag{4-4-31}$$

$$x_2(k) = x_2(0) - (kT)g \tag{4-4-32}$$

其中T为采样周期，g为重力加速度，为简化计算起见，暂设$T=1$，$g=1$（实际上重力加速度$g=9.8\text{m/s}^2$）。

相应的各相关函数为

$$\Phi_{x_1}(k,j)=E\left\{\left[x_1(0)+kx_2(0)-\frac{k^2}{2}\right]\left[x_1(0)+jx_2(0)-\frac{j^2}{2}\right]\right\} \tag{4-4-33}$$

$$\Phi_{x_1}(k,k)=E[x_1^2(k)]=E\left\{\left[x_1(0)+kx_2(0)-\frac{k^2}{2}\right]^2\right\} \tag{4-4-34}$$

$$\Phi_{x_2x_1}(k,j)=E\left\{[x_2(0)-k]\left[x_1(0)+jx_2(0)-\frac{j^2}{2}\right]\right\} \tag{4-4-35}$$

$$\Phi_{x_2}(k,k)=E[x_2^2(k)]=E\{[x_2(0)-k]^2\} \tag{4-4-36}$$

若给出初始条件，就可解出(4-4-33)～(4-4-36)式。初始值根据先验知识给定，若

$$\hat{x}_1(0)=h_0 \qquad \hat{x}_2(0)=v_0$$

这里把初始高度和速度看作是随机变量，令

$$\mathrm{var}[x_1(0)]=p_1 \qquad \mathrm{var}[x_2(0)]=p_2$$

则 $E[x_1^2(0)]=h_0^2+p_1 \qquad E[x_2^2(0)]=v_0^2+p_2$

因此，求得

$$\Phi_{x_1}(k,j)=\left(h_0+kv_0-\frac{k^2}{2}\right)\left(h_0+jv_0-\frac{j^2}{2}\right)+p_1+kjp_2 \tag{4-4-37}$$

$$\begin{aligned}\Phi_{x_1}(k,k)&=E[x_1^2(k)]\\&=\left(h_0+kv_0-\frac{k^2}{2}\right)^2+p_1+k^2p_2\end{aligned} \tag{4-4-38}$$

$$\Phi_{x_2x_1}(k,j)=(v_0-k)\left(h_0+jv_0-\frac{j^2}{2}\right)+jp_2 \tag{4-4-39}$$

$$\Phi_{x_2}(k,k)=(v_0-k)^2+p_2 \tag{4-4-40}$$

迄今，我们求出了所需要的各相关函数，进而就可求出在 k 时刻的滤波器的权系数。滤波器的权系数为 $H_1(k,i)$ 和 $H_2(k,i)$，$i=1, 2, \cdots, m$。

下面给出数字例子，给出初始值为

$$h_0=95\text{m}, \qquad v_0=0\text{m/sec}$$

$$p_1=10\text{m}^2, \qquad p_2=1(\text{m/sec})^2$$

计算结果如表4-4-1所示，表中给出了 $k=0$ 至 $k=6$ 各时刻落体的高度和速度的真值、观测值、估计值和估计的均方误差。

用维纳滤波方法求时变期望信号的估计是非常困难的，本例所以能够求解，是因为这里引入了描述落体运动规律的方程，从而求出了各相关函数。如果不引入落体的运动方程，是不能求出解答的。以后我们可以看到，用卡尔曼滤波方法来解决这个问题时将会简便得多，而且估计效果也较好。

表4-4-1 维纳滤波器对落体的高度和速度的估计

k	真值		高度测量值	估计值		均方误差	
	高度 $x_1(k)$	速度 $x_2(k)$	$z(k)$	高度 $\hat{x}_1(k)$	速度 $\hat{x}_1(k)$	高度 $p_1(k)$	速度 $p_2(k)$
0	100	0		95	0.00	10	1
1	99.5	−1.0	100	99.31	0.07	0.49	0.99
2	98.0	−2.0	97.9	99.45	−0.91	45.94	1.15
3	95.5	3.0	94.4	91.64	−1.89	11.66	1.31
4	92.0	−4.0	92.7	89.89	−2.88	7.80	1.49
5	87.5	−5.0	87.3	86.43	−3.86	2.89	1.68
6	82.0	−6.0	82.1	82.09	4.84	1.00	1.88

§4-5 广义维纳滤波

设输入测量矢量为$Z = X + W$，其中X为信号失量，W为噪声矢量，Z、X、W均为$N\times 1$维矢量，维数N很大，例如由一幅图像的象素组成，典型的大小为16×16。希望构造一个维纳滤波器，以便得到X的最小均方误差估计$\hat{X}_{l.m.s.}$。

因为期望信号X是一N维矢量，故相应的维纳滤波器的权系数构成一矩阵，维纳滤波器矩阵为

$$A_r = \Phi_{zx}^T \Phi_{zz}^{-1} \tag{4-5-1}$$

滤波估计为

$$\hat{X} = A_r Z = \Phi_{zx}^T \Phi_{zz}^{-1} Z \tag{4-5-2}$$

估计的均方误差为

$$P = \Phi_{xx} - \Phi_{zx}^T \Phi_{zz}^{-1} \Phi_{zx} \tag{4-5-3}$$

因为Z和X都是N维矢量，故A_r为$N\times N$维矩阵。由于N很大，滤波器结构将十分复杂，为了加以简化，以便滤波器容易实现，例如，若能够把A_r变换成一个对角线阵，那么，滤波的实施就变得容易多了。

为了简化滤波器矩阵A_r，引入变换矩阵T，T是一个适当选取的归一化正交矩阵，满足

$$TT^T = I, \quad T^{-1} = T^T \tag{4-5-4}$$

对输入矢量进行正交变换，得

$$F = TZ \tag{4-5-5}$$

令F通过滤波器A，滤波器的输出是ATZ，再经过逆变换，便得到估计$\hat{X}$

$$\hat{X} = T^{-1}ATZ = T^T ATZ \tag{4-5-6}$$

若根据最小均方误差准则来设计滤波器A，则A称为广义维纳滤波器，参见图4-5-1。

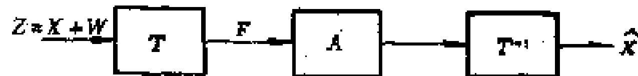


图4-5-1 广义维纳滤波结构

只要正交变换T选择适当，可以使滤波器矩阵A中的非零元素的数目大大减少，非零元素数目少意味着滤波器的权系数数目少，滤波器结构简单。若选择T是一种称之为卡南-洛也夫变换的最佳变换矩阵，还可以使A成为对角线阵，滤波器结构将变得最简单，最容易实施。

滤波器A的设计有两个途径，一是事先确定某种正交变换T，然后求出相应的广义维纳滤波器A，但这时A不是对角线阵。可供选择的正交变换T应当具有快速算法，例如1．DFT离散付立叶变换。2．WHT（沃尔什-阿达马变换）。3．DCT（离散余弦变换）。4．HT（哈尔变换）。5．ST（斜变换）等。我们把得到的滤波器矩阵A称为广义维纳滤波器。第二个途径是首先规定A是对角线阵，T是卡南-洛也夫变换，关于卡南-洛也夫变换将在以后解释，这时滤波器A称为对角线维纳滤波器。下面分别讨论广义维纳滤波器与对角线维纳滤波器的设计问题。

变换域协方差矩阵

设测量矢量Z是随机矢量，其均值为

$$E[Z]=\overline{Z}$$

那末，其方差阵定义为

$$\begin{aligned}P_z&=E[(Z-\overline{Z})(Z-\overline{Z})^T]\\&=E[ZZ^T]-\overline{Z}\,\overline{Z}^T\end{aligned}\tag{4-5-7}$$

输入矢量Z经过正交变换，得到变换域矢量F，记变换域的协方差阵为$\widetilde{P}_z$，F的均值为$\overline{F}$。

则
$$\begin{aligned}\widetilde{P}_z&=E[(F-\overline{F})(F-\overline{F})^T]\\&=E[FF^T]-\overline{F}\,\overline{F}^T\end{aligned}\tag{4-5-8}$$

把$TT^T=I$，$F=TZ$关系代入上式，得

$$\widetilde{P}_z=TP_zT^T=TP_zT^{-1}\tag{4-5-9}$$

这里利用了正交矩阵$T^{-1}=T^T$的性质，由(4-5-9)式可见，数据域协方差矩阵P_z与变换域协方差矩阵$\widetilde{P}_z$有相似关系，即符合矩阵相似性定义。

例4-5-1 设有一组数据矢量

$$Z_1=\begin{pmatrix}0\\2\end{pmatrix},\quad Z_2=\begin{pmatrix}1\\0\end{pmatrix},\quad Z_3=\begin{pmatrix}1\\1\end{pmatrix},\quad Z_4=\begin{pmatrix}1\\-1\end{pmatrix},\quad Z_5=\begin{pmatrix}2\\3\end{pmatrix}$$

是平稳矢量随机过程的采样，给定$T=\frac{1}{\sqrt{2}}\begin{pmatrix}1&1\\1&-1\end{pmatrix}$，求$P_z$和$\widetilde{P}_z$。

解：首先用平均法求出均值$\overline{Z}$的估计

$$\overline{Z}=\frac{1}{5}\sum_{i=1}^{5}Z_i=\begin{pmatrix}1\\1\end{pmatrix}$$

又 $$\overline{Z}\,\overline{Z}^T=\begin{pmatrix}1&1\\1&1\end{pmatrix}$$

$$E[ZZ^T]=\frac{1}{5}\left\{\begin{pmatrix}0\\2\end{pmatrix}[0,2]+\begin{pmatrix}1\\0\end{pmatrix}[1,0]\right.$$

$$\left.+\begin{pmatrix}1\\1\end{pmatrix}[1,\ 1]+\begin{pmatrix}1\\-1\end{pmatrix}[1,\ 1]+\begin{pmatrix}2\\3\end{pmatrix}[2,\ 3]\right\}$$

$$=\frac{1}{5}\begin{pmatrix}7&6\\6&15\end{pmatrix}$$

于是有 $$P_z=E[ZZ^T]-\overline{Z}\,\overline{Z}^T=\begin{pmatrix}0.4&0.2\\0.2&2\end{pmatrix}$$

$$\widetilde{P}_z=\frac{1}{2}\begin{pmatrix}1&1\\1&-1\end{pmatrix}\begin{pmatrix}0.4&0.2\\0.2&2\end{pmatrix}\begin{pmatrix}1&1\\1&-1\end{pmatrix}$$

$$=\begin{pmatrix}1.4&-0.8\\-0.8&2\end{pmatrix}$$

广义维纳滤波器矩阵A的导出

若数据域的先验统计知识为

$$\left.\begin{aligned}&E[W]=0,\ E[WW^T]=R,\ E[XW^T]=0\\ &E[ZZ^T]=\Phi_Z,\ E[XX^T]=\Phi_x\end{aligned}\right\}$$

自相关矩阵 (第二行)

则变换域的先验统计知识为

$$\widetilde{R}=TRT^T,\qquad \widetilde{\Phi}_x=T\Phi_xT^T,\qquad \widetilde{\Phi}_z=T\Phi_zT^T \tag{4-5-10}$$

输出估计的总均方误差为

$$J=E[\|\hat{X}-X\|]=E[(\hat{X}-X)^T(\hat{X}-X)]$$

$$=\mathrm{tr}E[(\hat{X}-X)(\hat{X}-X)^T] \tag{4-5-11}$$

把 $\hat{X}=T^{-1}ATZ=T^TATZ$代入上式，得

$$J=E[Z^TT^TATT^TATZ]-2E[Z^TT^TA^TTX]+E[X^TX] \tag{4-5-12}$$

利用关系$TT^T=1$，$F=TZ$代入上式得

$$J=E[F^TA^TAF]-2E[F^TA^TTX]-E[X^TX] \tag{4-5-13}$$

能使J达到最小时的A就是所求的广义维纳滤波器，它必定满足下述必要条件

$$\frac{\partial}{\partial A}J=0 \tag{4-5-14}$$

即有

$$E\left[\frac{\partial}{\partial A}(F^TA^TAF)\right]-2E\left[\frac{\partial}{\partial A}(F^TA^TTX)\right]+E\left[\frac{\partial}{\partial A}(X^TX)\right]=0 \qquad (4\text{-}5\text{-}15)$$

利用矩阵求导公式，我们有

$$\frac{\partial}{\partial A}(F^TA^TAF)=2AFF^T$$

$$\frac{\partial}{\partial A}(F^TA^TTX)=TXF^T$$

$$\frac{\partial}{\partial A}(X^TX)=0$$

(4-5-15)式成为

$$AE[FF^T]=TE[XF^T] \qquad (4\text{-}5\text{-}16)$$

把$F=TZ$关系代入，得

$$ATE[ZZ^T]T^T=TE[XZ^T]T^T \qquad (4\text{-}5\text{-}17)$$

因为
$$E[ZZ^T]=\Phi_z=\Phi_x+\Phi_w$$
$$E[XZ^T]=E[X(X+W)^T]=E[XX^T]=\Phi_x$$

于是有

$$AT\Phi_zT^T=T\Phi_xT^T \qquad (4\text{-}5\text{-}18)$$

因此，得出广义维纳滤波器矩阵A_0为

$$A_0=T\Phi_xT^T[T\Phi_zT^T]^{-1}$$
$$=\widetilde{\Phi}_x[\widetilde{\Phi}_z]^{-1}=\widetilde{\Phi}_x[\widetilde{\Phi}_x+\widetilde{\Phi}_w]^{-1} \qquad (4\text{-}5\text{-}19)$$

显然，(4-5-19)式与通常维纳滤波器的表达式类似，只是Φ_x、Φ_z都换成了变换域的$\widetilde{\Phi}_x$、$\widetilde{\Phi}_z$。

再把A_0写成另一形式

$$A_0=T\Phi_x\Phi_z^{-1}T^T=TA_rT^T \qquad (4\text{-}5\text{-}20)$$

其中$A_r=\Phi_x\Phi_z^{-1}$，即为通常意义下的维纳滤波器权矩阵，或称为维纳滤波器的冲击响应矩阵。

若X为零均值矢量，$\overline{X}=0$，令$P_x=\text{var}\,X$，$R=E[WW^T]$，则相关矩阵成为协方差矩阵。有

$$A_0=TP_x(P_x+R)^{-1}T^T \qquad (4\text{-}5\text{-}21)$$
$$A_r=P_x(P_x+R)^{-1} \qquad (4\text{-}5\text{-}22)$$

以变换域表示，则为

$$A_0=\widetilde{P}_x(\widetilde{P}_x+\widetilde{R})^{-1} \qquad (4\text{-}5\text{-}23)$$

其中 $\widetilde{P}_x=TP_xT^T$，$\widetilde{R}=TRT^T$。

最小均方误差J_{min}

因为均方误差$J=\mathrm{tr}\,E[(\hat{X}-X)(\hat{X}-X)^T]$，展开后得

$$J=\mathrm{tr}\ \{E[\hat{X}\hat{X}^T]-E[\hat{X}X^T]-E[X\hat{X}^T]+E[XX^T]\} \qquad (4\text{-}5\text{-}24)$$

由正交原理，得

$$E[(\hat{X}-X)\hat{X}^T]=0 \qquad 或 \qquad E[\hat{X}\hat{X}^T]=E[X\hat{X}^T]$$

代入关系式$\hat{X}=T^TATZ$ 及 $E[\hat{X}X^T]=T^TAT\Phi_x$，得

$$J=\mathrm{tr}[\Phi_x-T^TAT\Phi_x]$$

再利用关系式$A=A_0=T\Phi_x\Phi_z^{-1}T^T$，便得到最小均方误差$J_{min}$。

$$J_{min}=\mathrm{tr}[\Phi_x-\Phi_x\Phi_z^{-1}\Phi_x] \qquad (4\text{-}5\text{-}26)$$

用变换域表示，因为相似矩阵的迹相等，故有

$$J_{min}=\mathrm{tr}[\tilde{\Phi}_x-\tilde{\Phi}_x\tilde{\Phi}_z^{-1}\tilde{\Phi}_x] \qquad (4\text{-}5\text{-}27)$$

上述结果表明最小均方误差与所使用的正交变换T的形式无关。这一结论很有用处，我们可以任意选择正交变换T的形式，但不改变估计质量。

次最佳广义维纳滤波器

上面已得出结论，广义维纳滤波器的最小均方输出误差与正交变换T的种类无关。那末，最简单的形式是$T=1$，成为一般意义下的维纳滤波。但这时滤波器矩阵中非零元素太多，选择正交变换T希望滤波器矩阵中非零元素尽量少，同时还希望变换T有快速算法，以便能使滤波的实施容易。

若选定某种变换T，把广义维纳滤波器矩阵中数值较大的元素保留，而把数值较小的元素用零代替，只要不显著增大均方误差，这样做是允许的，这时，滤波器已不是最佳，故称为次最佳广义维纳滤波器。下面通过一例说明之。

例4-5-2　假定信号矢量是零均值的，其协方差矩阵为

$$\Phi_x=\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \cdots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{N-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \cdots & 1 \end{pmatrix}, \quad (0<\rho<1), \quad \Phi_w=\sigma_w^2 I_N$$

这类信号称为一阶马尔科夫序列，Φ_x称为陶布利兹矩阵，噪声矢量W是白噪声过程，其协方差矩阵为$R=\sigma_w^2 I_N$，选定 $N=16$，$\rho=0.9$，$\sigma_w^2=0.1$，并选择WHT（沃尔什-阿达马变换）作为正交变换T，计算出均方误差与滤波器非零元素数目之间的关系，同时为了比较起见，还计算$T=1$时的均方误差与滤波器非零元素数目之间的关系。

结果将表明，在均方误差保持相同的条件下，采用WHT时，滤波器中非零元素的数目显著减少，而且WHT滤波器的性能，在采用10个非零元素时，就很接近于保留全部256元素时的性能，见图4-5-2。

对角线维纳滤波器与KLT变换

一个合乎逻辑的提法是，选用什么样的正交变换T，才能使广义维纳滤波器成为对角线矩阵。

因为维纳滤波器矩阵　$A_r=P_x(P_x+R)^{-1}$　是一实对称矩阵，我们知道，一个实对称

矩阵必定能化为对角线矩阵。

对于实对称矩阵 B，其特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_N$，与特征值对应的特征矢量为 $q_1, q_2, \cdots, q_N$。记

$$
\left.\begin{aligned}
K &= [q_1, q_2, \cdots, q_N] \\
\Lambda &= \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{pmatrix}
\end{aligned}\right\} \qquad (4\text{-}5\text{-}28)
$$

则根据矩阵理论，存在下列重要关系，

$$
KBK^T = \Lambda \qquad (4\text{-}5\text{-}29)
$$

也就是说，利用正交变换矩阵 K，可以使实对称矩阵 B 变换成对角线阵 Λ。矩阵 K 称为卡南-洛也夫变换矩阵，简称KLT变换。

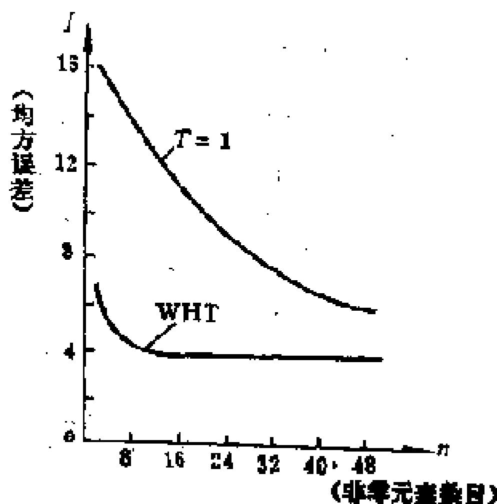


图4-5-2 滤波器A的非零元素数目与均方误差的关系

对于数据域矢量 $Z = X + W$，已知其维纳滤波器矩阵为

$$
A_r = P_x(P_x + P_w)^{-1}
$$

维纳滤波器输出为

$$
\hat{X} = A_r Z
$$

引入正交变换 T，$T^T T = 1$，则有

$$
\begin{aligned}
\hat{X} &= T^T T A_r T^T T Z \\
&= T^T (T A_r T^T)(T Z)
\end{aligned} \qquad (4\text{-}5\text{-}30)
$$

上式中 $TZ = F$ 为变换域矢量，TA_rT^T 为广义维纳滤波器，$T^T = T^{-1}$ 为逆变换，最后得到最优维纳滤波估计 $\hat{X}$。

现在选取正交变换 $T = KLT$，则有

$$
\begin{aligned}
\hat{X} &= K^T (K A_r K^T)(K Z) \\
&= K^T \Lambda (K Z)
\end{aligned} \qquad (4\text{-}5\text{-}31)
$$

此时广义维纳滤波器成为对角线阵 Λ。

例 给定

$$
P_x = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad (0 < \rho < 1) \text{ 和 } \quad P_w = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}
$$

求最佳对角线滤波器。

解：首先导出维纳滤波器为

$$
A_r = P_x(P_x + P_w)^{-1}
$$

$$
= \begin{pmatrix} \dfrac{2-\rho^2}{4-\rho^2} & \dfrac{\rho}{4-\rho^2} \\ \dfrac{\rho}{4-\rho^2} & \dfrac{2-\rho^2}{4-\rho^2} \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}
$$

其中 $\alpha=\frac{2-\rho^2}{4-\rho^2}$, $\beta=\frac{\rho}{4-\rho^2}$

再求A_r的特征值。特征方程为

$$|\lambda I-A_r|=0$$

或 $$\lambda^2-2\alpha\lambda+(\alpha^2+\beta^2)=0$$

解出特征值 $\lambda_1=\alpha+\beta$, $\lambda_2=\alpha-\beta$

相应的归一化特征矢量为

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} \qquad \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$$

于是得出KLT矩阵为

$$K=\frac{1}{2}\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}$$

从而可得最佳对角线滤波器为

$$\Lambda=KA_rK^{\mathrm{T}}$$

$$=\begin{bmatrix}\frac{2+\rho-\rho^2}{4-\rho^2} & 0\\ 0 & \frac{2+\rho-\rho^2}{4-\rho^2}\end{bmatrix}$$

考察最佳对角线滤波器，具有下列性质：

1. 它的输出均方误差最小，故称为最佳对角线滤波器。

2. 滤波器结构最简单，呈对角线矩阵，实施滤波时仅需进行标量乘法。

3. 由于 KLT 矩阵依赖于特定的矩阵 A_r，所以没有计算变换的快速算法，计算量很大。

4. 随着维数N的增大，求特征值变得很困难，不容易找到相应的KLT矩阵。

综上所述，KLT不适合工程实际应用，可是它是一切广义维纳滤波器中的最佳者，因此，常用它作为评价其它次最佳广义维纳滤波器的标准。

次最佳对角线滤波器

希望滤波器具有对角线形式，但又不希望采用KLT变换，而采用其它具有快速算法的某种已知正交变换，例如 DFT、DCT、WHT、ST等，人为地限定滤波器为对角线形式，一般说来，此时滤波器不再构成最佳维纳滤波器，其输出也不再具有最小均方误差。此滤波器称之为次最佳对角线滤波器。

设计次最佳对角线滤波器的方法如下：取滤波器矩阵A_d为对角线形式

$$A_d=\begin{pmatrix}a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{NN}\end{pmatrix} \tag{4-5-32}$$

令其中元素

$$a_{ii}=\frac{\widetilde{P}_x(i,i)}{\widetilde{P}_x(i,i)+\widetilde{R}(i,i)} \tag{4-5-33}$$

$\widetilde{P}_x(i,i)$ 为$\widetilde{P}_x$的对角线元素，$\widetilde{R}(i,i)$ 为$\widetilde{R}$ 的对角线元素，且已知

$$\widetilde{P}_x=TP_xT^T,\qquad \widetilde{R}=TRT^T$$

T为某种指定的正交变换。

由于这时滤波器是非最佳的，必须计算其均方误差，看看它偏离了最佳对角线滤波器的均方误差有多远。

滤波器A_d输出的均方误差为

$$\begin{aligned}J_d&=\mathrm{tr}[P_x-T^TA_dTP_x]\\&=\mathrm{tr}P_x-\mathrm{tr}(T^TA_dTP_x)\end{aligned} \tag{4-5-34}$$

由于 $\quad T^TT=1,\qquad \widetilde{P}_x=TP_xT^T$

故

$$P_x=T^T\widetilde{P}_xT \tag{4-5-35}$$

代入(4-5-34)式得

$$J_d=\mathrm{tr}P_x-\mathrm{tr}(T^TA_d\widetilde{P}_xT) \tag{4-5-36}$$

现在，因为 T是归一化的正交变换矩阵，故有

$$\mathrm{tr}P_x=\mathrm{tr}TP_xT^T=\mathrm{tr}\widetilde{P}_x \tag{4-5-37}$$

$$\mathrm{tr}(T^TA_d\widetilde{P}_xT)=\mathrm{tr}(A_d\widetilde{P}_x) \tag{4-5-38}$$

$$J_d=\mathrm{tr}\widetilde{P}_x-\mathrm{tr}(A_d\widetilde{P}_x) \tag{4-5-39}$$

把(4-5-33)式代入(4-5-39)中，得

$$J_d=\mathrm{tr}\widetilde{P}_x-\sum_{i=1}^{N}\frac{\widetilde{P}_x^2(i,i)}{\widetilde{P}_x(i,i)+\widetilde{R}(i,i)} \tag{4-5-40}$$

这个J_d就是所指定T的对角线滤波器的均方误差。由（4-5-26）式得出，最佳对角线滤波器的均方误差为

$$\begin{aligned}J_{min}&=\mathrm{tr}[P_x-P_xP_z^{-1}P_x]\\&=\sum_{i=1}^{N}\xi_i-\sum_{i=1}^{N}\xi_i^2/\eta_i\end{aligned} \tag{4-5-41}$$

其中ξ_i为P_x的特征值，η_i为P_x+R的特征值。

现在来比较各种次最佳对角线滤波器的性能。考察例 4-5-2，其他条件不变，取 $\rho=0.9$，$\sigma_n^2=1$，各种次最佳对角线滤波器在不同N值时的均方误差如表4-5-1所示。

KLT 的均方误差最小，以它为标准进行比较，由表中结果可见，DCT 最接近 KLT。但对于其他数据矢量则不一定，没有一定的规则可遵循，需要通过计算机模拟去寻找最接近KLT的具有快速算法的已知正交变换形式。

表4-5-1 各种对角线滤波器的均方误差

变换 \ N	2	4	8	16	32	64
KLT	5.9680	4.6640	4.0528	3.7696	3.6288	3.5584
DCT	5.9680	4.6720	4.0736	3.7894	3.6512	3.5712
DFT	5.9680	4.7424	4.3926	4.1472	3.9056	3.7120
WHT	5.9680	4.7072	4.2384	4.1312	4.1312	4.0944
HT	5.9680	4.7072	4.2400	4.1424	4.1312	4.1296

§4-6 连续时间维纳滤波

维纳最早提出的最佳滤波器是连续时间形式的，为了便于比较和理解，在这里对连续时间维纳滤波器加以扼要讨论。

采用的方法是对离散时间维纳滤波器采取样本加密并取极限，从而推导出连续时间维纳滤波器。

在连续时间域中，测量信号是 $z(t)$，期望信号是 $d(t)$，要得到在最小均方误差准则下的估计 $\hat{d}(t)$，假定 $z(t)$ 和 $d(t)$ 都是平稳过程，要求滤波器是时不变的和满足因果律的，即当 $t<0$ 时，滤波器的冲击响应 $h(t)=0$。

现在，从已知离散时间维纳滤波器出发，假定滤波器中延迟线的长度是无限的，利用所有过去的测量数据，即从 $t=-\infty$ 直到 $t=t_k$ 时刻的全部测量数据。

在 $t=t_k$ 时刻，测量矢量是

$$Z=[z_{-\infty},\ \cdots,\ z_l,\ \cdots,\ z_k]^T \qquad l\in(-\infty,\ k)$$

维纳滤波器的权矢量应当是

$$\boldsymbol{h}_k=\Phi_{zz}^{-1}\Phi_{zdk} \tag{4-6-1}$$

其中 d_k 为 k 时刻的期望信号，相应的维纳-霍夫方程为

$$\Phi_{zz}\boldsymbol{h}_k=\Phi_{zdk} \tag{4-6-2}$$

或写成时间函数形式有

$$\phi_{dz}(k,l)=\sum_{j=-\infty}^{k}h(k,j)\varphi_{zz}(j,l) \qquad -\infty<l\leqslant k \tag{4-6-3}$$

对于平稳过程有

$$\phi_{dz}(k-l)=\sum_{j=-\infty}^{k}h(k-j)\phi_{zz}(j-l) \tag{4-6-4}$$

现在令采样间隔 $\Delta t\to 0$，样本无限加密，离散时间变量代之以连续时间变量。

令 $\qquad k\sim t, \qquad l\sim\tau, \qquad j\sim\xi$

令和号换成积分号，则(4-6-4)式成为

$$\left.\begin{aligned}\phi_{dz}(t-\tau)=\int_{-\infty}^{t}h(t-\xi)\phi_{zz}(\xi-\tau)d\xi\\-\infty<\tau<t,\qquad -\infty<\xi<t\end{aligned}\right\}\tag{4-6-5}$$

再作一次变量置换，令

$$t-\tau=\eta,\qquad t-\xi=\lambda$$

其中　$0<\eta<\infty$，　　$0<\lambda<\infty$

则（4-6-5）式成为

$$\phi_{dz}(\eta)=\int_{0}^{\infty}h(\lambda)\phi_{zz}(\eta-\lambda)d\lambda\tag{4-6-6}$$

上式即为连续时间的维纳-霍夫方程，求解此方程则可求得连续时间维纳滤波器的冲击响应$h(t)$。

当$z(t)$为白色过程时，容易求得其解，代入$\phi(\eta-\lambda)=\delta(\eta-\lambda)$，则有

$$\begin{aligned}\phi_{dz}(\eta)&=\int_{0}^{\infty}h(\lambda)\delta(\eta-\lambda)d\lambda\\&=h(\eta),\qquad 0<\eta<\infty\end{aligned}\tag{4-6-7}$$

故维纳滤波器的冲击响应$h(t)$等于$d(t)$与$z(t)$的互相关函数$\phi_{dz}(t)$的$t>0$的那部分，即有

$$h(t)=\begin{cases}\phi_{dz}(t) & t\geqslant 0\\ 0 & t<0\end{cases}\tag{4-6-8}$$

故维纳滤波器在物理上是可实现的并符合因果律的。

对(4-6-8)式两边取拉氏变换，得

$$H(S)=[\Phi_{dz}(S)]_{+}\tag{4-6-9}$$

因为互功率谱$\Phi_{dz}(S)$的极点分布在S平面中的左右两边，可以用分部分式加以分开，

$$\Phi_{dz}(S)=[\Phi_{dz}(S)]_{+}+[\Phi_{dz}(S)]_{-}$$

其中$[\Phi_{dz}(S)]_{+}$只包括左半平面内的极点，$[\Phi_{dz}(S)]_{-}$只包含右半平面内的极点。

由于限定$h(t)=0$，$t<0$，故$H(S)$只取$[\Phi_{dz}(S)]_{+}$，那部分。$[\Phi_{dz}(S)]_{+}$称为满足因果律的传递函数。

当输入$z(t)$为非白色过程时，直接解维纳-霍夫方程是困难的，这时可用频率法求解。

令$z(t)$先通过-白化滤波器，使之变成白色过程$w(t)$，然后令$w(t)$再通过-最佳滤波器，如(5-6-9)式所示。

设输入信号功率谱$\Phi_{zz}(S)$为有理谱，其零点极点对称分布于S平面的左右两边，对$\Phi_{zz}(S)$分解，令$\Phi_{zz}(S)=\Phi_{zz}^{+}(S)\cdot\Phi_{zz}^{-}(S)$，其中$\Phi_{zz}^{-}(S)$只包含左半面的零点极点，$\Phi_{zz}^{-}(S)$只包含右半面的零点极点。

选取白化滤波器

$$H_{w}(S)=\frac{1}{\Phi_{zz}^{+}(S)}\tag{4-6-10}$$

根据线性系统理论，白化滤波器的输出功率谱为

$$\Phi_{zz}(S)H(S)H(-S)=1=\Phi_{w}(S)$$

显见，$w(t)$是白色过程。

再令白色过程$w(t)$通过最优滤波器$H_0(S)$，由(5-6-9)式可知

$$H_0(S)=[\Phi_{dw}(S)]_+ \tag{4-6-11}$$

其中$[\Phi_{dw}(S)]_+$为互功率谱$\Phi_{dw}(S)$中的因果律部分。现在来求$\Phi_{dw}(S)$。

由相关函数的定义有

$$\begin{aligned}\Phi_{dw}(\tau)&=E[d(t+\tau)w(t)]\\&=E\left[d(t+\tau)\int_{-\infty}^{\infty}h_w(\lambda)z(t-\lambda)d\lambda\right]\\&=\int_{-\infty}^{\infty}h_w(\lambda)\phi_{dz}(\tau+\lambda)d\lambda\end{aligned}$$

等式两边作双边拉氏变换，得互功率谱

$$\begin{aligned}\Phi_{dw}(S)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h_w(\lambda)\phi_{dz}(\tau+\lambda)e^{-S\tau}d\tau d\lambda\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h_w(\lambda)\phi_{dz}(\tau+\lambda)e^{-S(\tau+\lambda)}e^{S\lambda}d\tau d\lambda\\&=\left[\int_{-\infty}^{\infty}h_w(\lambda)e^{S\lambda}d\lambda\right]\Phi_{dz}(S)\\&=H_w(-S)\cdot\Phi_{dz}(S)=\frac{\Phi_{dz}(S)}{\Phi_{zz}^-(S)}\end{aligned} \tag{4-6-12}$$

取其因果律部分，得

$$H_0(S)=\left(\frac{\Phi_{dz}(S)}{\Phi_{zz}^-(S)}\right)_+$$

所以，整个维纳滤波器的传递函数为

$$H(S)=H_w(S)H_0(S)=\frac{1}{\Phi_{zz}^+(S)}\left(\frac{\Phi_{dz}(S)}{\Phi_{zz}^-(S)}\right)_+ \tag{4-6-13}$$

对$H(S)$取拉氏反变换，就得到冲击响应$h(t)$。滤波器的输出就是期望信号$d(t)$的估计

$$\widehat{d(t)}=\int_{-\infty}^{t}h(t-\xi)z(\xi)d\xi \tag{4-6-14}$$

输出均方误差为

$$\begin{aligned}P_{lmse}&=E[(d(t)-\widehat{d(t)})^2]\\&=E[(d(t)-\widehat{d(t)})d(t)]\\&=\phi_d(0)-\int_{-\infty}^{t}h(t-\xi)\phi_{dz}(t-\xi)d\xi\end{aligned} \tag{4-6-15}$$

这里利用了正交原理。作变量置换，令$\lambda=t-\xi$，于是有

$$P_{lmse}=\phi_d(0)-\int_0^{\infty}h(\lambda)\phi_{dz}(\lambda)d\lambda \tag{4-6-16}$$

例4-6-1 假设信号$x(t)$具有有理功率谱$\Phi_{xx}(\omega)=\dfrac{1}{1+\omega^2}$，白噪声的功率谱为

$\Phi_{vv}(\omega)=\dfrac{N_0}{2}$，信号与噪声不相关，试求维纳滤波器的冲击响应$h(t)$。

解：本题中期望信号是$x(t)$，接收信号为 $z(t)=x(t)+v(t)$

故有
$$\Phi_{zz}(S)=\Phi_{xx}(S)+\frac{N_0}{2} \tag{4-6-17}$$

$$\Phi_{xz}(S)=\Phi_{xx}(S) \tag{4-6-18}$$

代入（4-6-13）式中，得

$$H(S)=\frac{1}{\Phi_{zz}^{+}(S)}\left[\Phi_{zz}^{+}(S)-\frac{\Phi_{vv}(S)}{\Phi_{zz}^{-}(S)}\right]_{+}$$

$$=1-\frac{1}{\Phi_{zz}^{+}(S)}\left[\frac{N_0/2}{\Phi_{zz}^{-}(S)}\right]_{+} \tag{4-6-19}$$

由于
$$\Phi_{zz}(S)=\frac{1}{1-S^2}+\frac{N_0}{2}=\frac{1+\frac{N_0}{2}-\frac{N_0}{2}S^2}{1-S^2}$$

进行因子分解，得

$$\Phi_{zz}(S)=[\Phi_{zz}^{+}(S)]\cdot[\Phi_{zz}^{-}(S)] \tag{4-6-20}$$

$$\Phi_{zz}^{+}(S)=\frac{\sqrt{1+\frac{N_0}{2}}+\sqrt{\frac{N_0}{2}}S}{1+S} \tag{4-6-21}$$

$$\Phi_{zz}^{-}(S)=\frac{\sqrt{1+\frac{N_0}{2}}-\sqrt{\frac{N_0}{2}}S}{1-S} \tag{4-6-22}$$

又
$$\frac{\frac{N_0}{2}}{\Phi_{zz}^{-}(S)}=\sqrt{\frac{N_0}{2}}+\Phi_1(S) \tag{4-6-23}$$

由于$\Phi_1(S)$的极点在S平面的右边部分，故取

$$\left[\frac{N_0/2}{\Phi_{zz}(S)}\right]_{+}=\sqrt{\frac{N_0}{2}} \tag{4-6-24}$$

故得最优维纳滤波器的传递函数为

$$H(S)=1-\frac{\sqrt{\frac{N_0}{2}}(1+S)}{\sqrt{1+\frac{N_0}{2}}+\sqrt{\frac{N_0}{2}}S}=\frac{\sqrt{1+\frac{2}{N_0}}-1}{S+\sqrt{1+\frac{2}{N_0}}} \tag{4-6-25}$$

滤波器的冲击响应是

$$h(t)=\begin{cases}0 & t<0\\ \left(\sqrt{1+\frac{2}{N_0}}-1\right)e^{-\sqrt{1+\frac{2}{N_0}}\,t} & t\geqslant 0\end{cases} \tag{4-6-26}$$

考察一下此维纳滤波器的功率传递函数是有意义的，记 $a=\sqrt{1+\frac{2}{N_0}}$，则由(4-6-25)式，得

$$H(S)H(-S)=\frac{(a-1)^2}{a^2-S^2} \tag{4-6-27}$$

当$\frac{N_0}{2}$很小时，$a\to\infty$，这时滤波器的带宽很大，可以通过尽量多的信号功率。当$\frac{N_0}{2}$很大时，$a\approx 1$，这时滤波器带宽与有用信号带宽相同，使通过的噪声功率尽量少些。

例4-6-2 假设信号 $x(t)$的功率谱为 $\Phi_{xx}(S)=\frac{1}{1-S^2}$，噪声的功率谱为$\Phi_{vv}(S)=\frac{1}{4-S^2}$，信号与噪声不相关，求维纳滤波器的传递函数。

解：

$$\begin{aligned}\Phi_{zz}(S)&=\Phi_{xx}(S)+\Phi_{vv}(S)\\&=\frac{1}{1-S^2}+\frac{1}{4-S^2}\\&=\frac{5-2S^2}{(S^2-1)(S^2-4)}\\&=[\Phi_{zz}^{+}(S)]\cdot[\Phi_{zz}^{-}(S)]\end{aligned} \tag{4-6-28}$$

其中 $$\Phi_{zz}^{+}(S)=\frac{\sqrt{2}\ (S+\sqrt{2.5})}{(S+1)\ (S+2)}$$

$$\Phi_{zz}^{-}(S)=\frac{\sqrt{2}\ (S-\sqrt{2.5})}{(S-1)(S-}$$

故白化滤波器为

$$H_w(S)=\frac{(S+1)\ (S+2)}{\sqrt{2}(S+\sqrt{2.5})} \tag{4-6-29}$$

又

$$\begin{aligned}\frac{\Phi_{xx}(S)}{\Phi_{zz}^{-}(S)}&=\frac{S-2}{\sqrt{2}(S+1)(S-\sqrt{2.5})}\\&=\frac{0.822}{S+1}-\frac{0.115}{S-\sqrt{2.5}}\end{aligned}$$

取其因果律部分，得

$$H_0(S)=\left[\frac{\Phi_{xx}(S)}{\Phi_{zz}^{-}(S)}\right]_+=\frac{0.822}{S+1} \tag{4-6-30}$$

最终的维纳滤波器的传递函数为

$$\begin{aligned}H(S)&=H_w(S)H_0(S)\\&=\frac{(S+1)\ (S+2)}{\sqrt{2}\ (S+\sqrt{2.5})}\cdot\frac{0.822}{S+1}=\frac{0.582\ (S+2)}{S+\sqrt{2.5}}\end{aligned} \tag{4-6-31}$$

此滤波器容易用RC电路实现。

4-7 IIR型离散时间维纳滤波器

本章一开始，就强调离散时间维纳滤波器的结构是横向滤波器形式的，横向滤波器是有限冲击响应（FIR）的数字滤波器，故这一类维纳滤波器又称之为FIR型维纳滤波器，事实上还存在一种无限冲击响应（IIR）的维纳滤波器，其结构形式为IIR型数字滤波器，下面研究IIR型维纳滤波器。

在本章第一节中已研究过，对于平稳观测序列 $x_i=d_i+v_i$，$i=0, 1, \cdots, N$，矢量形式的维纳-霍夫方程为

$$R_{xd}=hR_{xx} \tag{4-7-1}$$

其中R_{xx}为观测序列的自相关矩阵，R_{xd}为观测序列与期望信号的互相关矢量，h 为维纳滤波器的权矢量。

把(4-7-1)式展开，得

$$R_{xd}(n)=\sum_{m=0}^{N}h(m)R_{xx}(n-m) \qquad n\in(0,\ N) \tag{4-7-2}$$

令$N\to\infty$，则有

$$R_{xd}(n)=\sum_{m=0}^{\infty}h(m)R_{xx}(n-m) \qquad n\in(0,\ \infty) \tag{4-7-3}$$

这时$h(m)$具有无限冲击响应。为了求解，首先考虑x_i为白色序列的情况，即有

$$R_{xx}(n-m)=\delta_{nm} \tag{4-7-4}$$

于是直接求得维纳滤波器的冲击响应为

$$h(n)=R_{xd}(n) \qquad n\in(0,\ \infty) \tag{4-7-5}$$

显然$h(n)$只在正时间轴上存在，滤波器是满足因果律的或物理可实现的。相应的传递函数为

$$H(z)=[\phi_{xd}(z)]^{+} \tag{4-7-6}$$

其中$\phi_{xd}(z)$为观测序列与期望信号的互功率谱，假定$\phi_{xd}(z)$是有理谱，加以分解，令

$$\phi_{xd}(z)=[\phi_{xd}(z)]^{+}+[\phi_{xd}(z)]^{-} \tag{4-7-7}$$

其中$[\phi_{xd}(z)]^{+}$只包含有单位圆内的极点，$[\phi_{xd}(z)]^{-}$只包含有单位圆上与单位圆外的极点。

如果x_i不是白色序列，可以把x_i先通过一个白化滤波器 $H_w(z)$，使之变成白色序列，白化滤波器$H_w(z)$的传递函数应是

$$H_w(z)=\frac{1}{\phi_{xx}^{+}(z)} \tag{4-7-8}$$

其中$\phi_{xx}(z)$为观测序列x_i的自动功率谱，假设$\phi_{xx}(z)$是有理谱，作谱因式分解，令

$$\phi_{xx}(z)=[\phi_{xx}^{+}(z)]\cdot[\phi_{xx}^{-}(z)] \tag{4-6-9}$$

其中$\phi_{xx}^{+}(z)$只含有单位圆内的零极点，而$\phi_{xx}^{-}(z)$只包含有单位圆外的零极点，令白化后的信号w_i再通过一最优滤波器$H_0(z)$，利用前面的结果，可知

$$H_0(z)=[\phi_{wd}(z)]_{+} \tag{4-7-10}$$

可以证明

$$\phi_{wd}(z) = H_w(z)\phi_{xd}(z) \tag{4-7-11}$$

而

$$H_w(z) = \frac{1}{\phi_{xx}^{-}(z)} \tag{4-7-12}$$

于是有

$$H_0(z) = \left[\frac{\phi_{xd}(z)}{\phi_{xx}^{-}(z)}\right]_+ \tag{4-7-13}$$

把$H_w(z)$与$H_0(z)$级联起来就得到整个维纳滤器波为

$$H(z) = H_w(z)H_0(z) = \frac{1}{\phi^{+}_{xx}(z)}\left[\frac{\phi_{xd}(z)}{\phi^{-}_{xx}(z)}\right]_+ \tag{4-7-14}$$

对$H(z)$作逆z变换就得到冲击响应$h(n)$。现在举一个例子。

例4-7-1 设有阶跃序列$d_i = A$，$i = 0$，1，2，$\cdots$，$x_i = 0$，$i < 0$，观测序列$x_i = d_i + v_i$，v_i为白噪声序列，已知$E[v_i] = 0$，$E[v_i^2] = \sigma^2$，求A的最优估计。

解： 出题意可知，观测序列的自功率谱为

$$\phi_{xx}(z) = \frac{A^2}{(z-1)(z^{-1}-1)} + \sigma^2$$

$$= \frac{\sigma^2 z_1^{-1}(z-z_1)(z^{-1}-z_1)}{(z-1)(z^{-1}-1)}$$

其中 $z_1 = \left(1 + \frac{A^2}{2\sigma^2}\right) - \sqrt{\left(1 + \frac{A^2}{2\sigma^2}\right)^2 - 1} < 1$

进行谱因式分解，得

$$\phi_{xx}^{+}(z) = \frac{\sigma}{\sqrt{z_1}} \cdot \frac{z - z_1}{z - 1}$$

$$\phi_{xx}^{-}(z) = \frac{\sigma}{\sqrt{z_1}} \cdot \frac{z^{-1} - z_1}{z^{-1} - 1}$$

又

$$\frac{\phi_{xd}(z)}{\phi_{xx}^{-}(z)} = \frac{A^2\sqrt{z_1}}{\sigma(z-1)(z^{-1}-z_1)}$$

由于$z_1 < 1$，显然上式在单位圆内无极点，因此，对于本例不存在IIR型维纳滤波器。

再来构造IIR型维纳滤波器，由(4-7-2)式及§4-1，得

$$(\sigma^2 + NA^2)\sum_{i=1}^{N} h_i = NA^2$$

$$h_1 = h_2 = \cdots = h_N = \frac{A^2}{NA^2 + \sigma^2}$$

对A的估计为

$$\hat{A} = \frac{1}{N + \frac{\sigma^2}{A^2}}\sum_{i=1}^{N} x_i$$

对于本例而言，FIR型 维纳滤波器是存在的而且是容易设计的。这也是FIR型维纳滤波器较之IIR型得到更为广泛应用的理由。

第五章　自适应维纳滤波

在上章中我们看到，维纳滤波是最小均方误差意义下的最优滤波，但要求期望信号是平稳的，并且要求知道输入信号序列的自相关矩阵 Φ_{zz} 和输入信号与期望信号的互相关矢量 Φ_{zd}。实际上，这些条件往往不满足，使得维纳滤波的应用直到很大限制。

所谓自适应维纳滤波，就是使横向滤波器的权系数按照某种算法自动地趋于最佳维纳权系数，并且当期望信号或输入信号统计量作缓慢变化时，只要计算速度跟得上，滤波器能自动地调整其权系数，并重新趋于新的最佳权系数。

自适应维纳滤波技术自七十年代以来发展很快，已发展成自适应噪声抵消，自适应波束形成，自适应线谱增强等新领域、并且还在继续发展之中。本章中将讨论自适应维纳滤波的基本理论及其主要应用。

§5-1　自适应维纳滤波原理

维纳滤波器的基本结构形式是一非递归型数字滤波器，滤波器的权系数按照均方误差最小准则达到最佳，因而滤波器的输出就是对期望信号的最小均方误差估计。

非递归型数字滤波器在工程上习惯地称为横向滤波器，横向滤波器加上其权系数的自动调节机构就构成了自适应维纳滤波器，其构造框图如图5-1-1所示，图5-1-2为其简化符号。

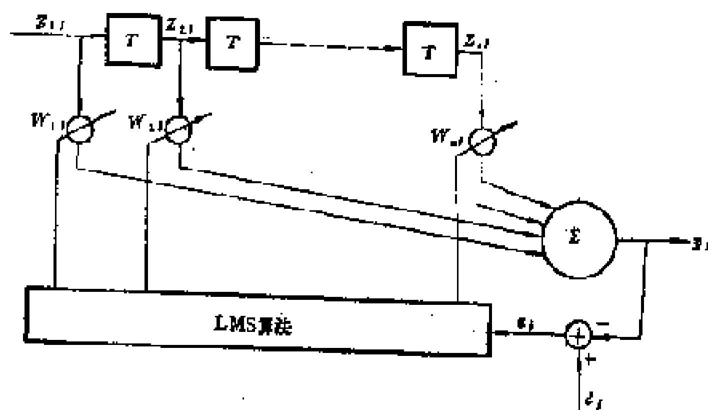


图5-1-1　自适应维纳滤波器

图中 Z_j 为 j 时刻上的输入数据；T 为抽头延迟时间；Z_{1j}，Z_{2j}，…，Z_{nj} 为 j 时刻延迟线上各抽头的信号；y_j 为 j 时刻上滤波器的输出；d_j 为 j 时刻的期望信号值；w_{1j}, w_{2j}，…，w_{nj} 为 j 时刻上滤波器的权系数。利用误差 $\varepsilon_j = d_j - y_j$，通过最小均方误差算法自动地去调节到最佳状态时，输出 y_j 就是 d_j 的最小均方误差估计 $\hat{d}_j$。这时横向滤波器的权矢量收敛于维纳滤波器最优权矢量，图5-1-2结构就称为自适应维纳滤波。

如果在起始时刻，滤波器的权系数不处于最佳状态，这时滤波器的输出不是的最佳估

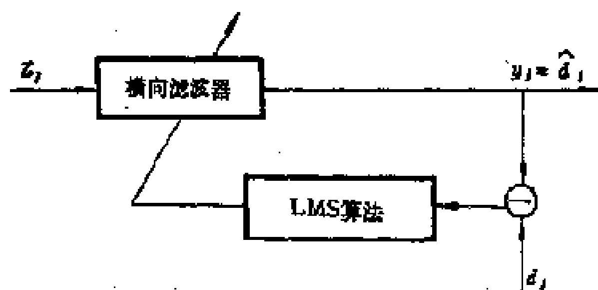


图5-1-2 自适应维纳滤波器的简化符号

计，滤波器就有较大的误差ε_j输出，通过lmse运算，使权系数逐步地调整，同时误差ε_j也逐步地减小，当达到稳态时，权系数接近最佳状态，滤波器的输出y_j就成为d_j的最佳估计，即有$y_j=d_j$。如果输入信号发生了新的变化，显然滤波器不处于最佳状态，误差ε_j增大，并再次去调整权系数，使滤波器重新进入新的最佳状态。整个系统的工作类似于一个自动调节机构，下面我们来研究误差ε_j是如何去调整权系数的。

§5-2 最速下降算法

定义j时刻滤波器的输入序列为Z_j

$$Z_j^T=[z_{1j},\ z_{2j},\ \cdots,\ z_{nj}] \tag{5-2-1}$$

j时刻滤波器的权系数为W_j

$$W_j^T=[w_{1j}, w_{2j}, \cdots, w_{nj}] \tag{5-2-2}$$

j时刻滤波器的输出为y_j

$$y_j=W_j^TZ_j=Z_j^TW_j \tag{5-2-3}$$

j时刻的输出误差为ε_j

$$\varepsilon_j=d_j-W_j^TZ_j=d_j-Z_j^TW_j \tag{5-2-4}$$

在j时刻，我们把权矢量W_j看成是一确定量，则均方误差$E[\varepsilon_j^2]$是W_j的函数，记输出均方误差为

$$E[\varepsilon_j^2]=J(W_j)=J_j \tag{5-2-5}$$

把(5-2-4)式代入(5-2-5)式，得

$$J(W_j)=E[d_j^2]-2E[d_jZ_j^T]W_j+W_j^TE[Z_jZ_j^T]W_j \tag{5-2-6}$$

记

$$\Phi_{zdj}=E[Z_jd_j^T],\quad \Phi_{dzj}=E[d_jZ_j^T] \tag{5-2-7}$$

Φ_{zdj}为j时刻上测量序列与期望信号的互相关矢量。

记

$$\Phi_{zzj}=E[Z_jZ_j^T] \tag{5-2-8}$$

Φ_{zzj}为j时刻上测量序列的自相关矩阵。

假定输入信号的统计特性是慢变化的，在一般短时间内，可以把输入信号看成是平稳的，因而在这段短时间内，可以把时标j省略，均方误差$J(W_j)$，仅决定于此时刻的权矢量W_j，故(5-2-6)式可改写成

$$J(W_j)=E[d_j^2]-2\Phi_{zd}^TW_j+W_j^T\Phi_{zz}W_j \tag{5-2-9}$$

因为权矢量W_j在自适应调节过程中是不断变化的，自适应滤波器实际上是一时变滤波器，故(5-2-9)式只在j时刻邻近的一小段时间内有效。在调整过程中，滤波器的输出也是非平稳的。

如果输入信号和期望信号是平稳的，那么滤波器经过一段过渡期之后进入稳态，此后，滤波器的输出将成为平稳的，而权矢量收敛于最优维纳权矢量。

下面来研究在此过渡期间权矢量W_j的变化规律，为简明起见，假设滤波器只有两个权w_1和w_2，展开(5-2-9)式，得

$$J(w_1,\ w_2)=E[d_j^2]-2[\Phi_{dz1}\ \ \Phi_{dz2}]\begin{bmatrix}w_1\\w_2\end{bmatrix}$$

$$+[w_1,\ w_2]\begin{bmatrix}\Phi_{z1z1} & \Phi_{z1z2}\\ \Phi_{z1z2} & \Phi_{z2z2}\end{bmatrix}\begin{bmatrix}w_1\\w_2\end{bmatrix}$$

$$=a_{11}w_1^2+2a_{12}w_1w_2+a_{22}w_2^2+2b_1w_1+2b_2w_2+c \qquad (5\text{-}2\text{-}10)$$

其中 $a_{11}=\Phi_{z_1z_1}=E[z_1^2]$，$a_{22}=E[z_2^2]$，$a_{12}=E[z_1z_2]$，

$c=E[d^2]$，$b_1=-E[dz_1]$，$b_2=-E[dz_2]$

显见，(5-2-10)式为椭圆方程式，每个J值就对应一个椭圆，$J(w_1,w_2)$连续变化就得到一个均方误差曲面，这是一个开口向上的椭圆抛物面，如图5-2-1所示。若滤波器有n个权，则均方误差曲面是一个$(n+1)$维空间中的超椭圆抛物面。

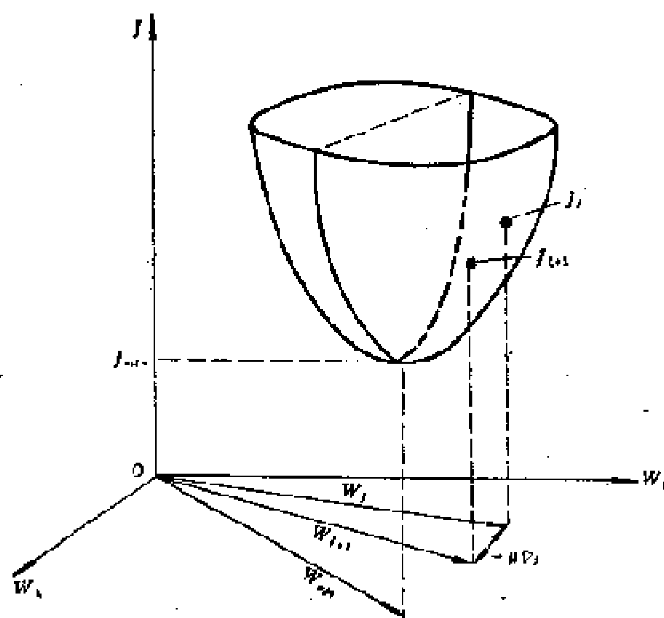


图5-2-1 均方误差曲面$J(W_1,\ W_2)$

在任一j时刻，权矢量W_j所对应的$J(W_j)$乃是该椭圆抛物面上的一个点，自适应调整权矢量就是要使$J(W_j)$沿着此椭圆抛物面下降，并尽快地到达最小值J_{min}，即到达此椭圆抛物面的底端。

为了尽快地到达最低点，最好是沿着$J(W)$减少率最大的方向去改变W_j，当$W=W_j$时，$J(W_j)$的梯度就是$J(W_j)$增加率最大的方向。$J(W_j)$的梯度的定义为

$$\nabla_j=\frac{\partial J}{\partial W_j}=\left[\frac{\partial J}{\partial w_{1j}},\ \frac{\partial J}{\partial w_{2j}},\ \cdots,\ \frac{\partial J}{\partial w_{nj}}\right]^T \qquad (5\text{-}2\text{-}11)$$

由(5-2-9)式得

$$\nabla_j=-2\Phi_{dz}+2\Phi_{zz}W_j \qquad (5\text{-}2\text{-}12)$$

梯度的负方向就是$J(W)$减少率最大的方向，我们使W_j沿着梯度的负方向改变一个量$\mu(-\nabla_j)$，其中μ是指定的一个小正常数，μ称为收敛因子，到$j+1$时刻，便得到权矢量W_{j+1}

$$W_{j+1}=W_j+\mu(-\nabla_j) \tag{5-2-13}$$

用(5-2-13)式的步骤去调整权矢量，就称为最速下降算法，μ 决定着每次调整的步长。

从几何直观上不难想象，只要 μ 值取得合适，W_j 经过若干次调整将逐步逼近并最终到达均方误差曲面的底端。如果 μ 取得太小，逼近速度会太慢；如果 μ 取得太大，可能出现不收敛现象。所以收敛因子是很重要的参数，它将直接影响到自适应滤波的效果。

当 J_j 已达到均方误差曲面的底端时，此刻的梯度 $\nabla_j=0$，权矢量将不再继续修正，也就是自适应滤波器到达稳态。此时，权矢量成为最佳权矢量 $W_j=W_{opt}$，由(5-2-12)式，得

$$W_{opt}=\Phi_{xx}^{-1}\Phi_{xd} \tag{5-2-14}$$

显见，(5-2-14)式就是熟知的维纳滤波器的权矢量表达式。这时自适应维纳滤波器就成了维纳滤波器。所以，利用最速下降算法，可以使自适应维纳滤波器收敛于维纳滤波器，当输入或期望信号变化时，滤波器的权矢量再次进行调整，并逐步到达新的平衡状态。

在第四章中我们知道，设计维纳滤波器时必须知道自相关矩阵 Φ_{xx} 和互相关矢量 Φ_{xd}，对于非平稳过程，这一点一般是难以做到的。而在自适应滤波器中，只要事先知道期望信号 d_j，就可能得到维纳滤波的结果。但是，仍然存在两个困难：一是如何求得梯度 ∇_j；一是如何得到期望信号 d_j。这两个问题暂且留在以后讨论。下面我们来研究最速下降算法所得结果的收敛性质。在这里，我们暂且假定期望信号 d_j 是已知的，梯度 ∇_j 是可以算出的。

1. 我们首先关心的问题是权矢量能否收敛于维纳解 W_{opt} 和在什么条件下才能收敛。

在 j 时刻，梯度 ∇_j 及均方误差 J_j 为

$$\left.\begin{aligned}\nabla_j&=-2\Phi_{dx}+2\Phi_{xx}W_j\\ J_j&=E[d_j^2]-2\Phi_{dx}^TW_j+W_j^T\Phi_{xx}W_j\end{aligned}\right\} \tag{5-2-15}$$

当滤波器为最佳状态时，我们有

$$\begin{aligned}W_{opt}&=\Phi_{xx}^{-1}\Phi_{dx}\\ J_{min}&=E[d_j^2]-2\Phi_{dx}^TW_{opt}+W_{opt}^T\Phi_{xx}\Phi_{xx}^{-1}\Phi_{dx}\\ &=E[d_j^2]-\Phi_{dx}^TW_{opt}\end{aligned} \tag{5-2-16}$$

由(5-2-16)式可知，要使均方误差收敛到 J_{min}，权矢量 W_j 应当收敛到维纳解 W_{opt}。记

$$V_j=W_j-W_{opt} \tag{5-2-17}$$

V_j 为权矢量的误差矢量，将 V_j 代入(5-2-15)式中，得

$$\nabla_j=2\Phi_{xx}V_j \tag{5-2-18}$$

$$J_j=J_{min}+V_j^T\Phi_{xx}V_j \tag{5-2-19}$$

已知在最速下降算法中采用如下迭代关系

$$W_{j+1}=W_j+\mu(-\nabla_j)$$

在上式两边减去 W_{opt} 得

$$\begin{aligned}V_{j+1}&=V_j-2\mu\Phi_{xx}V_j\\ &=(1-2\mu\Phi_{xx})V_j\end{aligned} \tag{5-2-20}$$

设矢量的初始值为 W_0，则 $V_0=W_0-W_{opt}$，在经 j 次迭代后，权的误差矢量为

$$V_j=(1-2\mu\Phi_{xx})^jV_0$$

显见，随着迭代次数增加，V_j应当趋于零，即有W_j趋近于W_{opt}，J_j趋近于J_{min}，由此得出权矢量W_j收敛于W_{opt}的条件可表达为

$$\lim_{j\to\infty}(1-2\mu\Phi_{xx})^j=0 \tag{5-2-22}$$

因为Φ_{xx}为实对称矩阵，必定具有各不相同的特征值，对Φ_{xx}作卡南-洛也夫变换得

$$L^{-1}\Phi_{xx}L=\Lambda \quad 或 \quad \Phi_{xx}=L\Lambda L^{-1} \tag{5-2-23}$$

其中Λ是由Φ_{xx}的特性值构成的对角阵，L称为模态矩阵，L的每一列是Φ_{xx}的归一化特征矢量(参见第四章)。

把(5-2-21)、(5-2-22)式改写成如下形式

$$1-2\mu\Phi_{xx}=L(1-2\mu\Lambda)L^{-1} \tag{5-2-24}$$

$$\lim_{j\to\infty}(1-2\mu\Phi_{xx})^j=\lim_{j\to\infty}L(1-2\mu\Lambda)^jL^{-1} \tag{5-2-25}$$

因而，权矢量收敛的等价条件为

$$\lim_{j\to\infty}(1-2\mu\Lambda)^j=0 \tag{5-2-26}$$

或写成等价式

$$\lim_{j\to\infty}\begin{pmatrix}(1-2\mu\lambda_1)^j & & & 0\\ & (1-2\mu\lambda_2)^j & & \\ & & \ddots & \\ 0 & & & (1-2\mu\lambda_n)^j\end{pmatrix}=0 \tag{5-2-27}$$

这要求 $$|1-2\mu\lambda_i|<1 \qquad i=1,2,\cdots,n \tag{5-2-28}$$

或 $$-1<1-2\lambda\mu_i<1$$

因为Φ_{xx}为对称正定矩阵，其全部特征值为正，故解出收敛条件为

$$\frac{1}{\lambda_i}>\mu>0 \qquad i=1,2,\cdots,n \tag{5-2-29}$$

令λ_{max}为Φ_{xx}的最大特征值，则收敛条件为

$$\frac{1}{\lambda_{max}}>\mu>0 \tag{5-2-30}$$

因此，只要收敛因子μ的取值满足(5-2-30)式的条件，最速下降算法自适应滤波器的权矢量就将收敛于维纳解W_{opt}，估计误差的均方值就将收敛于最小均方误差J_{min}，这时的滤波器也就成了维纳滤波器。

因为实际上我们并不知道Φ_{xx}的各个特征值和最大特征值λ_{max}，在实际应用中，可利用如下关系：

$$\lambda_{max}\leqslant\sum_{i=1}^{n}\lambda_i=\operatorname{tr}\Phi_{xx} \tag{5-2-31}$$

而 $$\operatorname{tr}\Phi_{xx}=\sum_{i=1}^{n}E[x_i^2]=总输入功率$$

总输入功率是可以测量到的。

因此，μ可按下式选取

$$\frac{1}{\mathrm{tr}\Phi_{zz}} \gg \mu > 0 \tag{5-2-32}$$

2．我们关心的第二个问题是收敛速度问题，即权矢量以多快速度向W_{opt}收敛。

权矢量的变化规律为

$$W_{j+1} = W_j + \mu(-\nabla_j)$$

其中 $\nabla_j = -2\Phi_{dz} + 2\Phi_{zz} W_j$

由(5-2-20)式已得知权的误差矢量V_{j+1}为

$$V_{j+1} = (1 - 2\mu\Phi_{zz})V_j$$

若Φ_{zz}为一对角线阵，其特征值为λ_1，λ_2，…，λ_n，则$(1-2\mu\Phi_{zz})$也为对角线阵，即有

$$\mu\Phi_{zz} = \begin{pmatrix} 1-2\mu\lambda_1 & & & 0 \\ & 1-2\mu\lambda_2 & & \\ & & \ddots & \\ 0 & & & 1-2\mu\lambda_n \end{pmatrix} \tag{5-2-33}$$

于是，(5-2-20)式可化为标量差分方程组，

$$v_i(j+1) = (1-2\mu\lambda_i)\, v_i(j) \qquad i = 1, 2, \cdots, n \tag{5-2-34}$$

这是一个等比级数衰减序列，可以近似地用一指数函数进行拟合，如图5-2-2所示。

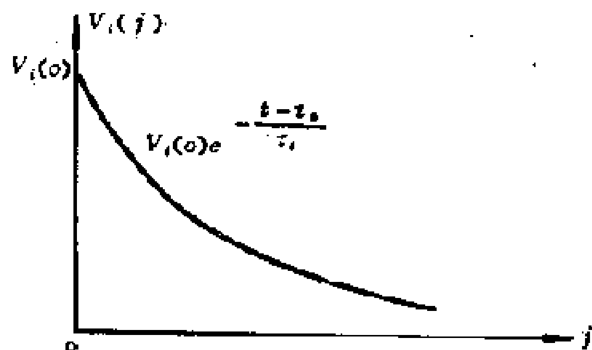


图5-2-2 权误差分量的变化规律

令 $v_i(j+1) = (1-2\mu\lambda_i)v_i(j)$

$$\approx e^{-\frac{T}{\tau}} v_i(j)$$

其中T采样周期，τ_i为待定时间常数。

由$e^{-\frac{T}{\tau_i}} = (1-2\mu\lambda_i)$得

$$\tau_i = \frac{-T}{\ln(1-2\mu\lambda_i)} \tag{5-2-35}$$

因为 $2\mu\lambda_i \ll 1$

$$\ln(1-2\mu\lambda_i) \approx -2\mu\lambda_i$$

故有

$$\tau_i \approx \frac{T}{2\mu\lambda_i} \tag{5-2-36}$$

若v_i的初始值为$v_i(0)$，则有

$$v_i(j) = v_i(0)e^{-\frac{jT}{\tau_i}} \tag{5-2-37}$$

可见权误差矢量的每一个分量按照各自的指数规律衰减。当μ增大时，时间常数τ_i变小，收敛速度变快。反之，当μ减少时，τ_i变大，收敛速度放慢。此外，τ_i还与Φ_{zz}的特征值λ_i有关，故各v_i的收敛速度不相同。

当Φ_{zz}不为对角线阵时，则$(1-2\mu\Phi_{zz})$也不是对角线阵，这时由(5-2-20)可看到，$v_i(j+1)$将是$v_1(j)$，$v_2(j)$，…，$v_n(j)$的线性组合，所以，$v_i(j+1)$将以n个指数衰减曲线之和的规律衰减。

当Φ_{zz}不是对角线阵时，这时需要对Φ_{zz}作卡南-洛也夫变换，先找出模态矩阵L，

我们有

$$V_{j+1}=L(1-2\mu\Lambda)L^{-1}V_j \tag{5-2-38}$$

记主坐标系中权的误差矢量为V_j^*，

$$V_j^*=L^{-1}V_j \tag{5-2-39}$$

则有

$$V_{j+1}^*=(1-2\mu\Lambda)V_j^* \tag{5-2-40}$$

若已知V_j^*的收敛规律，则不难求出V_j的收敛规律

$$V_j=LV_j^* \tag{5-2-41}$$

记V_j中的第i分量为$v_i(j)$，V_j^*中的第k分量为$v_k^*(j)$，L中第ik元素为l_{ik}，

则

$$v_i(j)=\sum_{k=1}^{n}l_{ik}v_k^*(j) \tag{5-2-42}$$

今已知$v_k^*(j)$按指数规律衰减，见(5-2-37)式。再由(5-2-42)式得知，$V(j)$中每一个分量均按n个指数衰减曲线之和进行衰减。

(5-2-40)式还可写成如下形式：

$$\begin{aligned}v_i(j)&=\sum_{k=1}^{n}v_k^*(0)l_{ik}(1-2\mu\lambda_k)^j\\&=\sum_{k=1}^{n}v_k^*(0)l_{ik}\exp\left[-\frac{jT}{\tau_k}\right]\end{aligned} \tag{5-2-43}$$

其中 $\tau_k=\dfrac{T}{2\mu\lambda_k}$，$\lambda_k$为$\Phi_{xx}$的第$k$个特征值。

又因为 $V_j=W_j-W_{opt}$

所以，由(2-4-3)式可以得出结论：权矢量W_j以多个几何衰减序列之和的规律衰减到W_{opt}。各衰减序列的几何比或时间常数，与收敛因子μ及对应的Φ_{xx}的特征值有关，当Φ_{xx}的各特征值相同时，W_j以简单几何衰减序列的规律收敛到W_{opt}。

3. 我们关心的第三个问题是均方误差J_j的收敛规律。由(5-2-19)式，得到在j时刻的均方误差为

$$\begin{aligned}J_j&=J_{min}+V_j^T\Phi_{xx}V_j\\&=J_{min}+V_j^{*T}L^T\Phi_{xx}LV_j^*\\&=J_{min}+V_j^{*T}\Lambda V_j^*\\&=J_{min}+\sum_{i=1}^{n}\lambda_i v_{ij}^{*2}\end{aligned} \tag{5-2-44}$$

再根据(5-2-34)及(5-2-37)式，把(5-2-44)式改写成

$$\begin{aligned}J_j&=J_{min}+\sum_{i=1}^{n}\lambda_i v_i^{*2}(0)(1-2\mu\lambda_i)^{2j}\\&=J_{min}+\sum_{i=1}^{n}\lambda_i v_i^{*2}(0)e^{-\frac{jT}{(\tau_i/2)}}\end{aligned} \tag{5-2-45}$$

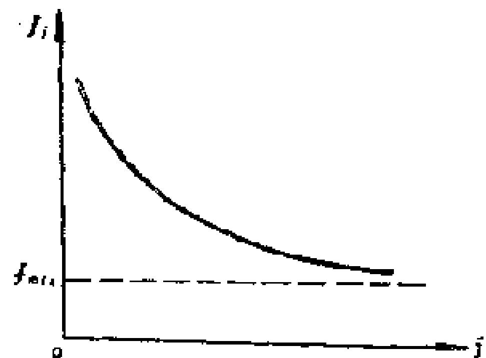


图5-2-3 最速下降算法的均方误差学习曲线

以J_j为纵坐标，以j为横坐标画出的曲线称为估计的均方误差学习曲线。均方误差学习

曲线与权矢量误差的学习曲线的形状相似，但其时间数要小一倍，且收敛终值是J_{min}，而不是零，如图5-2-3所示。

§5-3 威特罗(Widrow)算法

上节中对最速下降算法进行了详细研究，但是有两个问题尚未解决，故最速下降算法不能得到实际应用。一个是关于期望信号d_j，因为d_j正是我们需要去估计的量，不可能事先知道，如何获得d_j的问题留待下节中讨论；第二个是关于梯度∇_j的获得，这是本节中要讨论的问题。

最速下降算法的核心是要得到梯度∇_j。

$$\nabla_j=\frac{\partial J_j}{\partial W_j}=\frac{\partial}{\partial W_j}E[\varepsilon_j^2] \tag{5-3-1}$$

其中$E[\varepsilon_j^2]$是一个统计平均量，实际上是无法得到的。Widrow 建议用平方误差ε_j^2代替均方误差$E[\varepsilon_j^2]$。但对ε_j^2求梯度运算只能得到∇_j的一个估计$\hat{\nabla}_j$，而不是真正的梯度∇_j。

在j时刻上，梯度估计$\hat{\nabla}_j$定义为

$$\hat{\nabla}_j=\left[\frac{\partial\varepsilon_j^2}{\partial W_{1j}},\ \frac{\partial\varepsilon_j^2}{\partial W_{2j}},\ \cdots,\ \frac{\partial\varepsilon_j^2}{\partial W_{nj}}\right]^T \tag{5-3-2}$$

把$\varepsilon_j=d_j-\boldsymbol{Z}_j^T\boldsymbol{W}_j$代入上式得

$$\begin{aligned}\hat{\nabla}_j&=-2\varepsilon_j\boldsymbol{Z}_j=-2(d_j-\boldsymbol{Z}_j^T\boldsymbol{W}_j)\boldsymbol{Z}_j\\&=-2d_j\boldsymbol{Z}_j+2\boldsymbol{Z}_j^T\boldsymbol{W}_j\boldsymbol{Z}_j\end{aligned} \tag{5-3-3}$$

可以证明是$\hat{\nabla}_j$的无偏估计。对$\hat{\nabla}_j$取统计平均得

$$\begin{aligned}E[\hat{\nabla}_j]&=-2E[d_j\boldsymbol{Z}_j]+2E[\boldsymbol{Z}_j\boldsymbol{Z}_j^T]\boldsymbol{W}_j\\&=-2\Phi_{dz}+2\Phi_{zz}\boldsymbol{W}_j=\nabla_j\end{aligned}$$

显见，$\hat{\nabla}_j$是∇_j的无偏估计。

用$\hat{\nabla}_j$代替∇_j，就得到权矢量$\hat{\boldsymbol{W}}_j$的迭代算法，$\hat{\boldsymbol{W}}_j$是$\boldsymbol{W}_j$的一个估计。

$$\begin{aligned}\hat{\boldsymbol{W}}_{j+1}&=\hat{\boldsymbol{W}}_j+\mu(-\hat{\nabla}_j)\\&=\hat{\boldsymbol{W}}_j+2\mu\varepsilon_j\boldsymbol{Z}_j\end{aligned} \tag{5-3-4}$$

(5-3-4)式所表示的算法就称为Widrow算法，误差ε_j和测量$\boldsymbol{Z}_j$都可以实时得到，给出初值$\hat{\boldsymbol{W}}_0$之后。即可算出$\hat{\boldsymbol{W}}_j$，这里仍假定期望信号d_j是已知的。

注意到在最速下降算法中，∇_j与$\boldsymbol{W}_j$之间存在确定的关系，即$\nabla_j=-2\Phi_{dz}+2\Phi_{zz}\boldsymbol{W}_j$，只要测量信号与期望信号是平稳的，$\Phi_{dz}$与$\Phi_{zz}$就是确定的，故$\nabla_j$为确定量。但是在威特罗算法中，由(5-3-2)式可见，ε_j和$\boldsymbol{Z}_j$都是随机量，故权矢量$\hat{\boldsymbol{W}}_j$也是随机量，而且是非平稳随机矢量时间序列。所以，由威特罗算法得到的权矢量将不同于由最速下降算法得到的权矢量。在此情况下，自适应滤波器的收敛条件是什么？按照什么规律收敛？均方误差的收敛规律又是什么？这些问题是我们希望了解的，下面就来研究之。

1、关于权矢量的均值$E[\hat{\boldsymbol{W}}_j]$的收敛条件。因为在威特罗算法中，$\hat{\boldsymbol{W}}_j$是随机序列，讨论$\hat{\boldsymbol{W}}_j$的收敛问题是没有意义的，故我们讨论$E[\hat{\boldsymbol{W}}_j]$的收敛条件。

在j次迭代上，我们有$\varepsilon_j=d_j-\boldsymbol{Z}_j^T\boldsymbol{W}_j$，梯度估计$\hat{\nabla}_j$被定义为

$$\hat{\nabla}_j = \begin{pmatrix} \dfrac{\partial \varepsilon_j^2}{\partial w_{1j}} \\ \dfrac{\partial \varepsilon_j^2}{\partial w_{2j}} \\ \vdots \\ \dfrac{\partial \varepsilon_j^2}{\partial w_{nj}} \end{pmatrix} = \begin{pmatrix} 2\varepsilon_j \dfrac{\partial \varepsilon_j}{\partial w_{1j}} \\ 2\varepsilon_j \dfrac{\partial \varepsilon_j}{\partial w_{2j}} \\ \vdots \\ 2\varepsilon_j \dfrac{\partial \varepsilon_j}{\partial w_{nj}} \end{pmatrix} = -2\varepsilon_j \boldsymbol{Z}_j \tag{5-3-5}$$

或

$$\hat{\nabla}_j = -2(d_j - \boldsymbol{Z}_j^T \boldsymbol{W}_j)\boldsymbol{Z}_j = -2d_j \boldsymbol{Z}_j + 2\boldsymbol{Z}_j \boldsymbol{Z}_j^T \boldsymbol{W}_j \tag{5-3-6}$$

把(5-3-6)式代入(5-3-4)式中，得

$$\hat{\boldsymbol{W}}_{j+1} = \hat{\boldsymbol{W}}_j + 2\mu\varepsilon_j \boldsymbol{Z}_j = \hat{\boldsymbol{W}}_j + 2\mu d_j \boldsymbol{Z}_j - 2\mu \boldsymbol{Z}_j \boldsymbol{Z}_j^T \boldsymbol{W}_j \tag{5-3-7}$$

对(5-3-7)式取统计平均得

$$E[\hat{\boldsymbol{W}}_{j+1}] = E[\hat{\boldsymbol{W}}_j] + 2\mu E[d_j \boldsymbol{Z}_j] - 2\mu E[\boldsymbol{Z}_j \boldsymbol{Z}_j^T \boldsymbol{W}_j] \tag{5-3-8}$$

在(5-3-8)式中含有三阶混合矩，分析是困难的，为了简化讨论，我们假设 $(\boldsymbol{Z}_j \boldsymbol{Z}_j^T)$ 与 $\boldsymbol{W}_j$ 的互协方差矩阵为零，即

$$E\{[\boldsymbol{Z}_j \boldsymbol{Z}_j^T - E(\boldsymbol{Z}_j \boldsymbol{Z}_j^T)][\hat{\boldsymbol{W}}_j - E\hat{\boldsymbol{W}}_j]\} = 0 \tag{5-3-9}$$

把(5-3-9)式展开，得

$$E[\boldsymbol{Z}_j \boldsymbol{Z}_j^T \hat{\boldsymbol{W}}_j] - E[\boldsymbol{Z}_j \boldsymbol{Z}_j^T] \cdot E[\hat{\boldsymbol{W}}_j] = 0$$

即有

$$E[\boldsymbol{Z}_j \boldsymbol{Z}_j^T \hat{\boldsymbol{W}}_j] = E[\boldsymbol{Z}_j \boldsymbol{Z}_j^T] \cdot E[\hat{\boldsymbol{W}}_j] \tag{5-3-10}$$

(5-3-9)式成立的条件是根据 $\hat{\boldsymbol{W}}_j = \hat{\boldsymbol{W}}_{j-1} + 2\mu\varepsilon_{j-1}\boldsymbol{Z}_{j-1}$ 中 $\hat{\boldsymbol{W}}_j$ 与 $\boldsymbol{Z}_{j-1}$，$\boldsymbol{Z}_{j-2}, \cdots, \boldsymbol{Z}_0$，有关，而与 $\boldsymbol{Z}_j$ 无关。把(5-3-10)式代入(5-3-8)式中，得

$$E[\hat{\boldsymbol{W}}_{j+1}] = E[\hat{\boldsymbol{W}}_j] + 2\mu\Phi_{dz} - 2\mu\Phi_{zz} E[\hat{\boldsymbol{W}}_j] \tag{5-3-11}$$

记权误差矢量的均值为 $E[\hat{\boldsymbol{V}}_j]$

$$E[\hat{\boldsymbol{V}}_j] = E[\hat{\boldsymbol{W}}_j - \hat{\boldsymbol{W}}_{opt}] \tag{5-3-12}$$

$\boldsymbol{W}_{opt}$ 为最佳权矢量，$\boldsymbol{W}_{opt} = \Phi_{zz}^{-1}\Phi_{dz}$。

把(5-3-11)式改写成如下形式：

$$E[\hat{\boldsymbol{W}}_{j+1}] = E[\hat{\boldsymbol{W}}_j] - 2\mu\Phi_{zz}(E[\hat{\boldsymbol{W}}_j] - \boldsymbol{W}_{opt}) \tag{5-3-13}$$

上式两边同时减去 $\boldsymbol{W}_{opt}$，则得权误差矢量的递推式

$$E[\hat{\boldsymbol{V}}_{j+1}] = E[\hat{\boldsymbol{V}}_j] - 2\mu\Phi_{zz} E[\hat{\boldsymbol{V}}_j] = (1 - 2\mu\Phi_{zz})E[\hat{\boldsymbol{V}}_j] \tag{5-3-14}$$

显见，威特罗算法中 $E[\boldsymbol{V}_j]$ 的变化规律与最速下降算法中 $\boldsymbol{V}_j$ 的变化规律相同。类似于最速下降算法中的讨论，我们可以推知，$E[\boldsymbol{W}_j]$ 收敛于 $\boldsymbol{W}_{opt}$ 的条件是

$$\lim_{j\to\infty}(1 - 2\mu\Lambda)^j = 0$$

$$或 \quad \frac{1}{\lambda_{max}} > \mu > 0 \tag{5-3-15}$$

其中 Λ 为 Φ_{zz} 的特征值构成的对角线阵，λ_{max} 为 Φ_{zz} 的最大特征值。由于 λ_{max} 为未知，利用以下关系

$$\lambda_{max} \leqslant \sum_{i=1}^{n} \lambda_i = \mathrm{tr}\Phi_{zz} = \sum_{i=1}^{n} E[Z_{ij}^2]$$

$$= 总输入功率$$

其中总输入功率是容易测得的，在实用上，收敛因子μ 可取值为

$$\frac{1}{\mathrm{tr}\Phi_{xx}} \gg \mu > 0 \tag{5-3-16}$$

在满足(5-3-15)条件下，$E[\hat{W}_j]$将收敛于维纳解W_{opt}，权矢量$\hat{W}_j$是由$E[\hat{W}_j]$和权噪声矢量所组成。

2. 关于$E[\hat{W}_j]$的收敛规律。类似于最速下降算法中的讨论，我们可以推知，$E[\hat{W}_j]$将以几个等比级数之和衰减，各个等比级数的比值是$r_i = 1 - 2\mu\lambda_i$，$i = 1, 2, \cdots, n$。当 Φ_{xx} 各特征值相同时，$E[\hat{W}_j]$即以单一等比级数形式衰减。

我们也可以用指数曲线去拟合等比级数，各指数曲线的时间常数为

$$\tau_i = \frac{T}{2\mu\lambda_i} \tag{5-3-17}$$

3. 输出均方误差的稳态性质。已知输出均方误差为

$$J_j = E[\varepsilon_j^2] = E[d_j^2] - 2\Phi_{dj}^T W_j + W_j^T \Phi_{xxj} W_j \tag{5-3-18}$$

问当迭代步数 $j \to \infty$时，J_j 的收敛性质如何？

现在来研究这个问题。利用关系式

$$W_{opt} = \Phi_{xx}^{-1}\Phi_{xd}$$

$$J_{min} = E[d^2(t)] - W_{opt}^T \Phi_{xd}$$

不难把(5-3-18)式改写成

$$J_j = J_{min} + (W_j - W_{opt})^T \Phi_{xxj}(W_j - W_{opt}) \tag{5-3-19}$$

在自适应迭代过程中，因W_j只是以均值趋于W_{opt}，所以，稳态均方根差J_{ss}不趋于J_{min}，

$$J_{ss} = \lim_{j\to\infty} J_j \neq J_{min} \tag{5-3-20}$$

而是多出了一项由权噪声引起的误差功率。

均方误差J_j 的变化规律称为自适应过程的学习曲线，通常把$(J_{ss} - J_{min})$称为超量均方误差，典型的威特罗算法学习曲线如图5-3-1所示。



图5-3-1 威特罗算法学习曲线

我们定义失调量M

$$M = \frac{J_{ss} - J_{min}}{J_{min}} \tag{5-3-21}$$

失调量M是描述自适应维纳滤波器性能的一个重要参数，它描述了由于采用估计梯度而付出的代价。

在实际系统中，真实梯度是不知道的，必需采用估计梯度，因而必然使输出均方误差增大。在工程应用中，M不大于10%已认为满意。

因为最小均方误差J_{min}无法知道，故不能直接应用(5-3-21)式，通常用下式估算失调量

$$M = \frac{J_{ss} - J_{min}}{J_{min}} \approx \frac{\mu \mathrm{tr}\Phi_{xx}}{1 - \mu \mathrm{tr}\Phi_{xx}} \tag{5-3-22}$$

其中 $\mathrm{tr}\,\phi_{zz}=\sum_{i=1}^{n}E[z_i^2]$ = 总输入功率，总输入功率是容易测量到的，根据选定的收敛因子μ，就可算出失调M。

已经证明，μ 控制着$E[\widehat{W}_j]$的收敛速度，选择 $0<\mu<\frac{1}{\lambda_{max}}$，可保证$E[\widehat{W}_j]$的收敛到$W_{opt}$。现在还知道，$\mu$ 还控制着失调M的大小，为了使M保持在较小的数值上，应选取$\mu\,\mathrm{tr}\Phi_{zz}\ll 1$，或 $0<\mu\ll\frac{1}{\mathrm{tr}\Phi_{zz}}$，这与前面讨论权矢量收敛条件时所得结果相同。

已经知道，增大收敛因子μ 可以加快收敛速度，但由(5-3-22)式知道，增大μ 同时使失调M增大。可见，在自适应调整过程中，这二者是互助矛盾的，在工程上，常常采用折中办法来选取μ 值，或根据规定的失调选取μ 值。

其他自适应算法

在Widrow算法中(又称WLMS算法)，权矢量的迭代关系式为

$$W_{j+1}=W_j+2\mu\varepsilon_jZ_j$$

这里需要进行ε_j和Z_j的乘法运算，如有m个权就需要m个乘法器，给硬件实现带来了困难，为了便于工程实现，提出了下述简化算法。

1. 限幅LMS算法，又称CLMS算法。其权矢量的迭代关系为

$$W_{j+1}=W_j+2\mu\varepsilon_j\mathrm{sgn}[Z_j] \tag{5-3-23}$$

其中sgn为符号函数，$\mathrm{sgn}[Z_j]=\begin{cases}1 & x\geqslant 0\\ -1 & x<0\end{cases}$

2. 符号LMS算法，又称SLMS算法，其权矢量的迭代关系为

$$W_{j+1}=W_j+2\mu\,\mathrm{sgn}[\varepsilon_j]\mathrm{sgn}[Z_j] \tag{5-3-24}$$

若把μ 取成2的负整数幂，则权的运算完全不用乘法器。

上述两种算法的性能要作严格的数学分析比较困难，但近似分析和实验结果表明，限幅的结果使权矢量收敛的稳定性较好，但收敛速度有所下降。

3. 还有一种格律费史(Griffiths)算法，其权矢量的迭代关系为

$$W_{j+1}=W_j+2\mu[\widehat{\Phi}_{zd}-Z_jZ_j^TW_j] \tag{5-3-25}$$

其中$\widehat{\Phi}_{zd}$为观测矢量与期望信号的互相关矢量的估计，由

$$\begin{aligned}W_{j+1}&=W_j-\mu\nabla_j\\&=W_j+2\mu(\Phi_{zd}+\Phi_{zz}W_j)\end{aligned}$$

用$Z_jZ_j^T$ 代替相关矩阵$\Phi_{zz}=E[ZZ^T]$，再用估计量$\widehat{\Phi}_{zd}$代替Φ_{zd}，就得到格律费史算法。

§5-4 自适应维纳滤波的应用

在以上各节讨论中，都假定期望信号d_j已经知道，实际上d_j是要估计的量，不可能事先知道。所以，要使得自适应维纳滤波器能够获得实际应用，必须先解决d_j的获得问题。由于在不同的工程应用中期望信号的获取方式是不同的，所以，将结合应用实例来说明期望信号的获取方法。

一、自适应维纳预测器

已往知道，在维纳滤波器中，j 时刻的输出是期望信号d_j的最小均方误差估计$\hat{d}_j$。

$$y_j = \hat{d}_j = \Phi_{d_z}^{T} \Phi_{zz}^{-1} Z_j = W_{opt}^{T} \cdot Z_j$$

其中Z_j为 j 时刻的测量值矢量

$$Z_j = [z_{1j}, \ z_{2j}, \ \cdots, \ z_{nj}]^T$$

参照图5-4-1，若取期望信号d_j为$j+\Delta$ 时刻的输入采样值 $z_{(j+\Delta)}$，即令 $d_j = z_{(j+\Delta)}$，那未，输出y_j将是$z_{(j+\Delta)}$的估计，也就是对z_j作出了Δ步长的预报。但是，由于 $z_{(j+\Delta)}$ 是未知的，故互相关矢量 Φ_{dz} 也无法知道，故无法求出最佳权矢量W_{opt}，也就是不能够设计出需要的维纳预测器。然而，若采用自适应维纳滤波器的结构，却可以组成自适应维纳预测器。

自适应预测器的构造如图5-4-1所示。

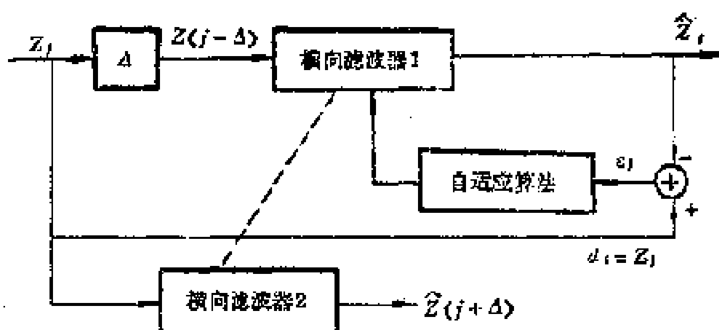


图5-4-1　自适应维纳预测器

对照图5-1-1中自适应维纳滤波器结构，可见图5-4-1的上部乃是一自适应维纳滤波器。期望信号即为 j 时刻输入采样z_j，把z_j延迟 Δ步后输入横向滤波器 1，通过自适应算法自动调整横向滤波器 1 的权矢量，其输出即为z_j的估计$\hat{z}_j$。横向滤波器 2 的权矢量与横向滤波器 1 的权矢量同步复制，其输入为z_j，那么,输出就是z_j的 Δ步预报估计 $\hat{z}_{(j+\Delta)}$。

二、自适应干扰抵消器

在接收信号中，除了有用信号$s(t)$之外，常常混有来源已知的干扰$n_0(t)$，这种干扰与 $s(t)$的频谱重叠，因此不能用通常的滤波方法抑制，然而采用自适应滤波方法却可以加以抑制。

图5-4-2给出了自适应干扰抵消器的基本结构。下面研究自适应干扰抵消器是如何抑制干扰的。

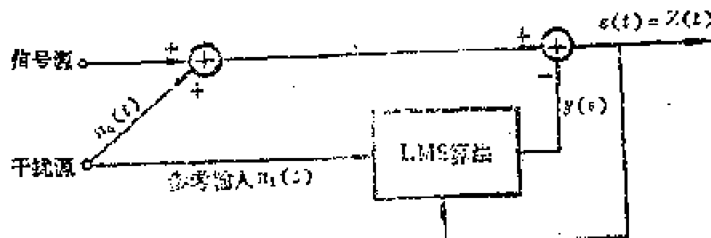


图5-4-2　自适应干扰抵消器

设接收信号是有用信号 $s(t)$加上干扰$n_0(t)$，称之为原输入，再把干扰$n_1(t)$加到横向滤波器的输入端，称为参考输入。横向滤波器的输出是 $y(t)$，减法器的输出是 $z(t) = s(t) + n(t) \quad y(t)$，$z(t)$同时通过自适应算法去调整横向滤波器的权。

把图5-4-2与图5-1-1中自适应维纳滤波器相比较。显见，期望信号是$s(t)+n_0(t)$，输入是$n_1(t)$。因为输出是$z(t)=s(t)+n_0(t)-y(t)$，如果$y_0(t)$能够逼近$n_0(t)$，那么，输出$z(t)$就逼近有用信号$s(t)$，也就是说$n_0(t)$被抵消了，这就是自适应干扰抵消器名称的由来。

$n_1(t)$和$n_0(t)$来自同一干扰源，但经过不同的传输路径，而且传输路径的信道特性是缓慢变化的，这是最常见的情况，因而$n_0(t)$与$n_1(t)$都是时变的，它们不可能相同，但必定保持着一定程度的相关性。再假设有用信号$s(t)$与$n_0(t)$、$n_1(t)$不相关。下面研究自适应干扰抵消器的工作原理。已知

$$z(t)=s(t)+n_0(t)-y(t) \tag{5-4-1}$$

用离散时标j换去连续时标t得

$$z_j=s_j+n_{0j}-y_j$$

两边取平方得

$$z_j^2=s_j^2+(n_{0j}-y_j)^2+2s_j(n_{0j}-y_j) \tag{5-4-2}$$

两边取统计平均，得

$$E[z_j^2]=E[s_j^2]+E[(n_{0j}-y_j)^2]+2E[s_j(n_{0j}-y_j)] \tag{5-4-3}$$

已经规定s_j与n_{0j}不相关，而y_j与n_{0j}通过横向滤波器（线性系统）所产生，故s_j与y_j也不相关，于是有

$$E[z_j^2]=E[s_j^2]+E[(n_{0j}-y_j)^2] \tag{5-4-4}$$

自适应运算是按均方误差最小化进行的，也就是使$E[z_j^2]$趋向最小化。在自适应调整过程中，y_j受到调整，因为信号s_j与n_{1j}不相关，故信号s_j不受影响，所以，$E[z_j^2]$趋向最小化也就是$E[(n_{0j}-y_j)^2]$趋向最小化。因此，横向滤波器的输出y_j将成为n_{0j}的最小均方误差估计。由于，$z_j-s_j=n_{0j}-y_j$，显然，输出z_j就成为有用信号s_j的最小均方误差估计。

根据以上分析，还可以得出以下几点推论：

1． 若原输入和参考输入中含有不相关的成分m_{0j}和m_{1j}，则输出信噪比会降低。

根据对自适应维纳滤波器的分析可知，只有期望信号d_j与输入矢量x_j相关时，滤波器才能作出d_j的估计。对于自适应干扰抵消器而言，只有当n_{1j}与n_{0j}相关时，y_j才能成为干扰n_{0j}的估计$y_j=\hat{n}_{0j}$，使得原输入中的干扰n_{0j}被抵消，从而抵消器的输出z_j成为有用信号s_j的估计，即得到$z_j=s_j$。对于不相关的背景噪声，显然抵消器是无能为力的，因此，抵消器的输出信噪比将会下降。

2． 若参考输入中含有有用信号成分，将会使输出信噪比下降。

因为两个输入中的有用信号相关，故y_j中亦含有s_j的成分，在相减运算时，有用信号也被抵消，因而输出信噪比下降。所以，为了保证良好的干扰抵消效果，参考输入中不能含有s_j的成分。

3． 影响抵消器效能的因素还在于采用了估计梯度而不是真实梯度。由于采用的是估计梯度，使得输出y_j中含有失调噪声，这种失调噪声的存在使得对n_{0j}的估计误差增大，从而降低了抵消器的干扰抵消效果。

4． 影响抵消器效能的第四个因素是横向滤波器的长度有限。从理论上讲，理想维纳滤波器的冲击响应在时间轴上向正负两端无限延伸，但实际上横向滤波器的长度是有限的，用有限维的权矢量去逼近无限维的维纳滤波器的冲击响应，必然存在误差，任何实际的自

适应维纳滤波器不可能达到理想维纳滤波器的性能。在自适应抵消器中，y_j也不可能等于n_{0j}，也就是说，原输入中的干扰不可能被完全抵消。

自适应干扰抵消器实例

下面介绍一个已用硬件实现的自适应干扰抵消器的实例，图5-4-3是它的结构框图，图5-4-4是其中一个权运算电路。

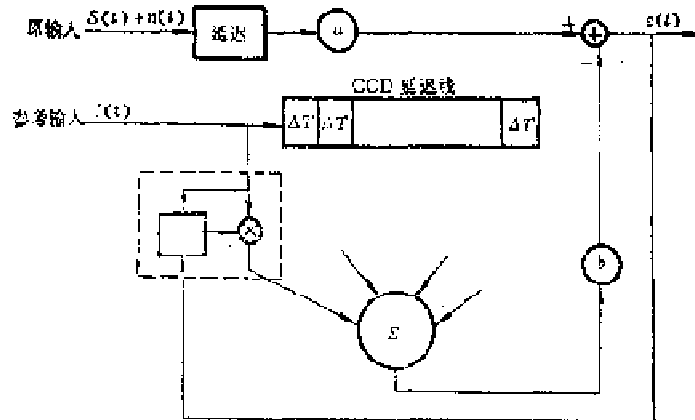


图5-4-3 自适应干扰抵消器实例框图

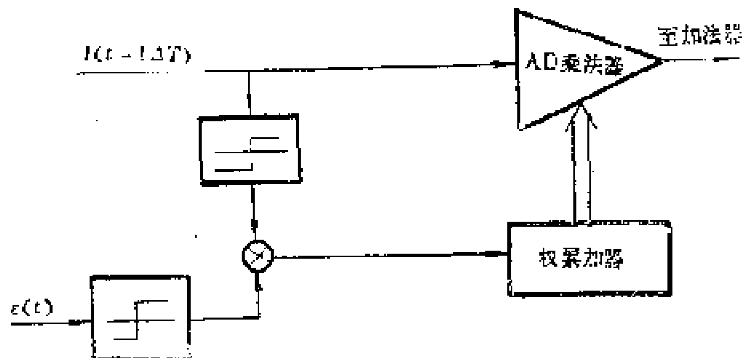


图5-4-4 图中的一个权运算电路

横向滤波器采用CCD器件作为延迟线，CCD各节的输出是时间量化的采样保持平顶信号，参考输入取自干扰源$I(t)$，期望信号取自原输入，即$s(t)+n(t)=d(t)$，其中$s(t)$为有用信号，噪声$n(t)$与$I(t)$来自同一干扰源。期望信号$d(t)$为时间连续信号，而CCD延迟线的抽头输出时间离散信号。

自适应运算的核心是完成权矢量的迭代运算

$$W_{j+1}=W_j+2\mu\varepsilon_j I_j \qquad (5\text{-}4\text{-}5)$$

设延迟线的长度为L节，则W_j为L维矢量，故需要L个权运算电路。图5-4-4为其中的第l个权运算电路，它完成一个权值的迭代运算，现在来研究这个运算电路是如何完成运算的。

第l节抽头的输出是$I(t-l\Delta T)$，误差信号是$\varepsilon(t)$，它们的幅度都是模拟量，因为难以得到良好稳定性的模拟乘法器，同时为了简化设备，故采用图5-4-4所示的方案，即采用SLMS算法。这里把$I(t-l\Delta T)$和$\varepsilon(t)$都加以极性化。

令

$$\operatorname{sgn}(x)=\begin{cases}1 & x>0\\ 0 & x=0\\ -1 & x<0\end{cases} \qquad (5\text{-}4\text{-}6)$$

于是在 j 时刻，我们有 $\mathrm{sng}\,I(t-l\cdot\Delta T)=I_{j-l}$，$\mathrm{sgn}\,\varepsilon(t)=\varepsilon_j$，在极性乘法器中完成了二者的相乘 $\varepsilon_j I_{j-l}$。乘积 $\varepsilon_j I_{j-l}$ 送入权累加器，在权累加器中，原来已存储有上一时刻的权值 $W_j(l)$，于是累加器的权变成

$$w_{j+1}(l)=w_j(l)+\varepsilon_j I_{j-l} \qquad l=0,1,\cdots,L-1 \tag{5-4-7}$$

在 j 时刻，横向滤波器的输出为 y_j

$$y_j=\sum_{l=0}^{L-1}w_j(l)I_{j-l} \tag{5-4-8}$$

误差信号 $\varepsilon_j=d_j-y_j$，由减法器输出，改变 a 和 b 的比值就相当于改变收敛因子 μ 。

I_{j-l} 是CCD 的抽头输出，其振幅是模拟量，而 $w_j(l)$ 是累加寄存器输出，它是个二进制数字量，为得到二者的乘积，这里采用了AD乘法器，L 个AD乘法器的输出相加就得出了横向滤波器输出 y_j，y_j 实际上是时间连续信号 $y(t)$。

误差 $\varepsilon(t)=d(t)-y(t)$，一方面送到权运算电路，一方面又作为抵消器的输出。根据对抵消器工作原理的分析得知，当抵消器进入稳态后，即有 $y(t)=\hat{n}(t)$ 和 $\varepsilon(t)=\hat{d}(t)-y(t)=\hat{s}(t)$。如果抵消器的电路设计良好，性能稳定，那么，原输入中的干扰 $n(t)$ 将大部分被抵消，输出信噪比将有显著提高。

图5-4-5为实验电路的实测结果，图 a 为其中一个权的平均收敛曲线，权累加寄存器中的数是二进制数，用MC1408DAC使之变成模拟量，以 1 ms为重复周期，观测到的平均变化规律；图 b 为抵消器的学习曲线，以 1 ms为重复周期，对输出 $\varepsilon(t)$ 取平方，在示波器上观测到输出均方值 $E[\varepsilon(t)^2]$。

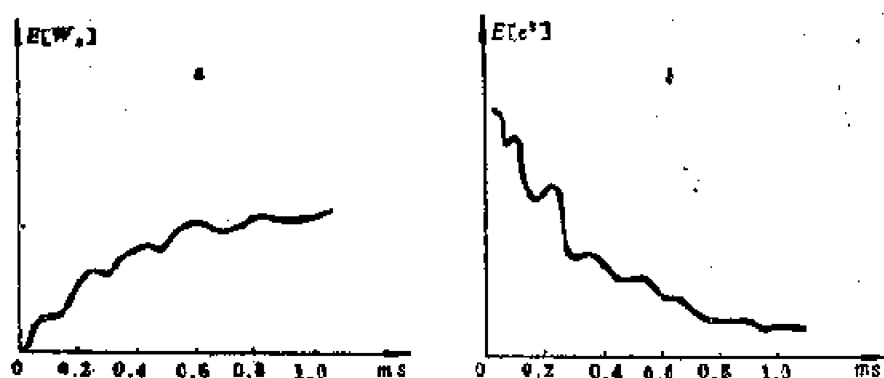


图5-4-5 自适应干扰抵消器的实验结果

a—$E[w]$收敛规律 b—$E[\varepsilon^2]$收敛规律

三、自适应线谱增强器

假设在噪声背景中有若干个正弦波，噪声背景很强，把正弦信号完全淹没，而且噪声是有色的，其相关时间达到 k 个采样周期，这些正弦波的频率是未知的，且是慢变化的，我们希望能够检测到这些正弦波并跟踪其频率变化。用一般方法是很难实现上述要求的，这时需要使用自适应线谱增强器。

图5-4-6是自适应线谱增强器的原理图，下面研究其工作原理，图5-4-6实际上是一个 Δ 步自适应预测器。

图中期望输入取自观测序列，参考输入是观测序列的 Δ 步延迟，令 $\Delta>k$，因为已假定 d_j 中的背景噪声的相关时间小于 Δ 步，因此，参考输入中的噪声 $n_{j-\Delta}$ 与 d_j 中的噪声

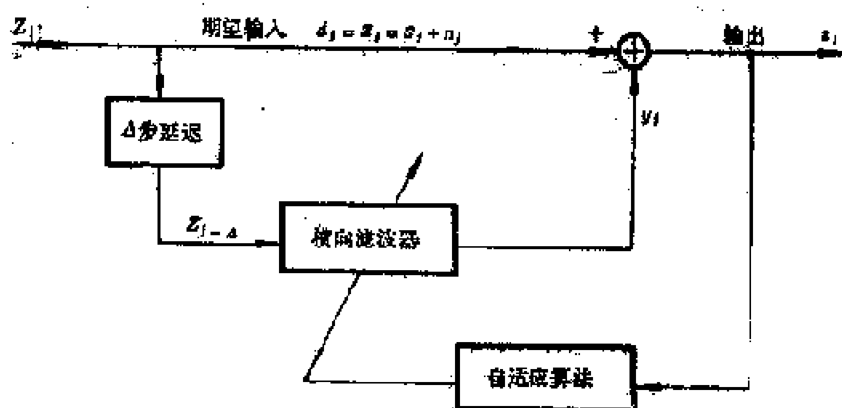


图5-4-6 自适应线谱增强器原理图

n_j不相关。不相关的$n_{j-\Delta}$不可能利用来构成对n_j的合理估计，$n_{j-\Delta}$通过横向滤波器只能使输出预报误差均方值增大，因此，为了保证输出均方误差最小化，自适应算法必然要对$n_{j-\Delta}$尽可能地加以抑制。另一方面，延时Δ只对正弦波引入相移，在参考输入端与原输入端之间仍保持良好的相关性，因此，自适应算法使y_j中只含有正弦波，并在减法器中被良好地抵消。


图5-4-7 线谱增强器中横向滤波器的传递函数

总之，横向滤波器抑制噪声通过，而让诸正弦波通过。从频率域观点来看，横向滤波器的传递函数在诸正弦频率上形成窄峰，而在其它地方下陷，当正弦波频率发生漂移时，这些窄峰将随之移动，于是保持着高的输出信噪比。自适应线谱增强器的传递函数如图5-4-7所示。

最后对横向滤波器的输出y_j作付立叶分析，就能检测出输入信号z_j中的线谱成分，当线谱的频率缓慢变化时，线谱增强器能自动地进行跟踪。

四、自适应陷波器

对自适应陷波器的传递函数的要求与对自适应线谱增强器的要求正好相反，希望在正弦波频率处形成尖的深谷，而在其它频率上保持平坦，这些正弦波频率是未知的，且是慢变化的，我们希望抑制了这些干扰正弦波。对于这个问题，用固定的陷波器是无能为力的，而用自适应方法是可以解决的。

本节只分析单频陷波器的性能，图5-4-8为单频陷波器的原理图。因为对于单频波来说，只有频率和相位两个未知参量（振幅无关紧要），所以只需要用两个可调权。参考输入取自干扰源$z(t)=A\cos(\omega_0 t+\phi)$，期望信号$d(t)$等于干扰正弦波加上其它任何类信号。

单频自适应陷波器的工作原理如下：首先把参考输入分解为两路正交信号，于是有

$$\left.\begin{aligned} z_{1j}&=A\cos(\omega_0 jT+\phi)\\ z_{2j}&=A\sin(\omega_0 jT+\phi)\end{aligned}\right\}\qquad(5\text{-}4\text{-}9)$$

其中，T为采样周期

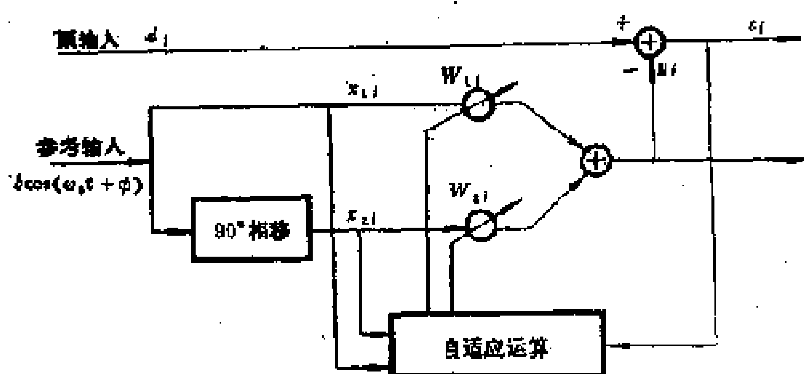


图5-4-8　单频自适应陷波器

令　权系数的迭代关系为

$$\left.\begin{aligned} w_{1(j+1)} &= w_{1j} + 2\mu\varepsilon_j x_j \\ w_{2(j+1)} &= w_{2j} + 2\mu\varepsilon_j x_j \end{aligned}\right\} \qquad (5\text{-}4\text{-}10)$$

为了计算单频陷波器的传递函数，画出其信号流程图，如图5-4-9所示。

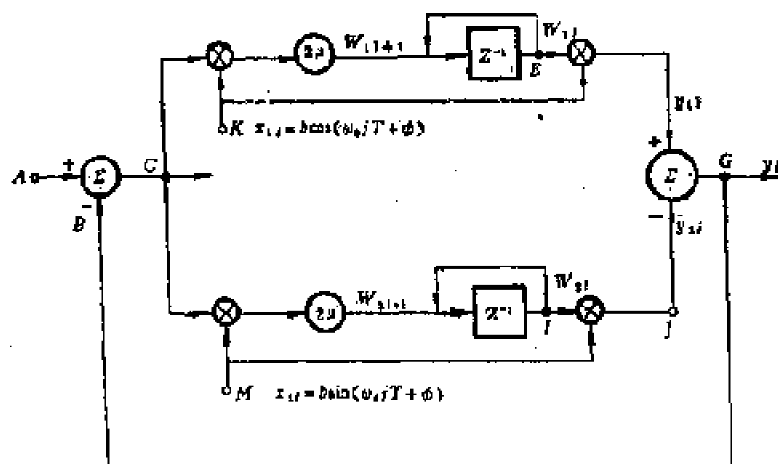


图5-4-9　单频陷波器的信号流程图

分析的第一步是获得从ε_j(C点)到y_j(G点)的开环传递函数。打开B点的反馈路径，同时在C点加入一单位冲击

$$\varepsilon_j = \delta_{j-k} \qquad \delta_{j-k} = \begin{cases} 1 & k = j \\ 0 & k \neq j \end{cases} \qquad (5\text{-}4\text{-}11)$$

于是在D点的响应为

$$\varepsilon_j x_{1j} = \begin{cases} A\cos(\omega_0 kT + \phi) & k = j \\ 0 & k \neq j \end{cases} \qquad (5\text{-}4\text{-}12)$$

从D点到E点的传递函数为$\dfrac{2\mu}{Z+1}$，其冲击响应为$2\mu u_{j-1}$，其中$u_j = \begin{cases} 0, & j < 0 \\ 1, & j \geqslant 0 \end{cases}$是一个单位阶跃函数。

对$\varepsilon_j x_{1j}$与$2\mu u_{j-1}$求乘积，就得出E点的响应为

$$w_{1j} = 2\mu A\cos(\omega_0 kT + \phi) \qquad j \geqslant k+1 \qquad (5\text{-}4\text{-}13)$$

F 点的输出响应为

$$y_{1j} = 2\mu A^2\cos(\omega_0 jT+\phi)\cos(\omega_0 kT+\phi) \tag{5-4-14}$$

其中 $j \geqslant k+1$。同理，可以得到 I 点的响应

$$y_{2j} = 2\mu A^2\sin(\omega_0 jT+\phi)\sin(\omega_0 kT+\phi) \tag{5-4-15}$$

由此，得到 G 点的响应为

$$y_j = 2\mu A^2\cos[\omega_0 T(j-k)]u_{j-k-1} \tag{5-4-16}$$

(5-4-16)式是 $(j-k)$ 的函数，所以，这是一个时不变的冲击响应。不失一般性，令 $k=0$，则有

$$y_j = h(j) = 2\mu A^2\cos(\omega_0 jT)\cdot u(j-1) \tag{5-4-17}$$

所以，从 C 点到 G 点的传递函数为

$$\begin{aligned} G(Z) &= 2\mu A^2\left[\frac{Z(Z-\cos\omega_0 T)}{Z^2-2Z\cos\omega_0 T+1}-1\right] \\ &= \frac{2\mu A^2[Z\cos\omega_0 T+1]}{Z^2-2Z\cos\omega_0 T+1} \end{aligned} \tag{5-4-18}$$

令 $T=\dfrac{2\pi}{\Omega}$，则

$$G(Z) = \frac{2\mu A^2[Z\cos(2\pi\omega_0\Omega^{-1})-1]}{Z^2-2Z\cos(2\pi\omega_0\Omega^{-1})+1} \tag{5-4-19}$$

再接通 G 点到 B 点的反馈路径，则从输入点 A 到 C 点的闭环传递函数为

$$H(Z) = \frac{Z^2-2Z\cos(2\pi\omega_0\Omega^{-1})+1}{Z^2-2(1-\mu A^2)\cos(2\pi\omega_0\Omega^{-1})Z+1-2\mu A^2} \tag{5-4-20}$$

令分子等于零，得 $H(Z)$ 的零点为

$$\begin{aligned} Z_0 &= \cos(2\pi\omega_0\Omega^{-1}) \pm \sqrt{\cos^2(2\pi\omega_0\Omega^{-1})-1} \\ &= \exp(\pm j2\pi\omega_0\Omega^{-1}) \end{aligned} \tag{5-4-21}$$

令分母等于零，得 $H(Z)$ 的极点为

$$\begin{aligned} Z_p &= (1-\mu A^2)\cos(2\pi\omega_0\Omega^{-1}) \\ &\quad \pm j[(1-2\mu A^2)+(1-\mu A^2)^2\cos^2(2\pi\omega_0\Omega^{-1})]^{\frac{1}{2}} \end{aligned} \tag{5-4-22}$$

Z_p 至圆心的距离为

$$\rho = (1-2\mu A^2)^{\frac{1}{2}} \approx 1-\mu A^2 \tag{5-4-23}$$

极点的幅角为

$$\begin{aligned} \Psi &= \cos^{-1}[(1-\mu A^2)(1-2\mu A^2)^{-\frac{1}{2}}\cos(2\pi\omega_0\Omega^{-1})] \\ &\approx \pm 2\pi\omega_0\Omega^{-1} \end{aligned} \tag{5-4-24}$$

根据 $H(Z)$ 的零点极点在 Z 平面上的位置，容易画出幅频响应 $H(\omega)$，如图(5-4-10)所示。

因为零点位于单位圆上，故 $H(\omega)$ 在 $\omega=\omega_0$ 时的响应等于零。因此，自适应陷波器对正弦干扰具有良好的抑制作用。又极点非常靠近零点，相距近似等于 μA^2，因此，$H(\omega)$ 的功率带宽大约是

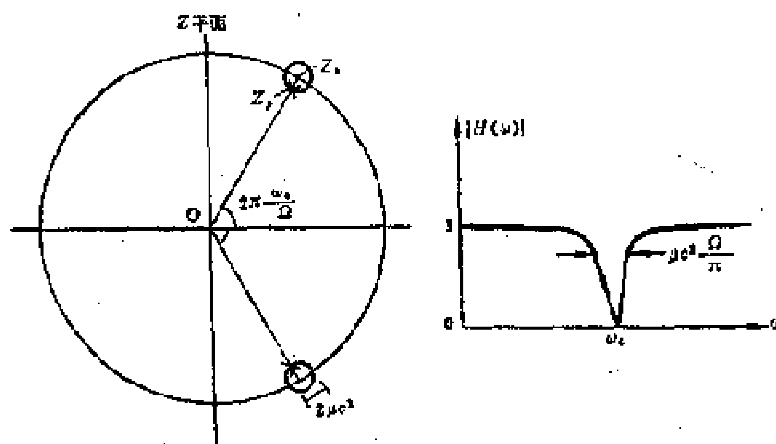


图5-4-10 单频陷波器的特性
a—Z平面上零点极点　　b—幅频特性

$$\mathrm{BW}\approx 2\mu A^2\frac{\Omega}{2\pi}=\frac{\mu A^2\Omega}{\pi} \tag{5-4-25}$$

陷波器的Q值等于中心频率与带宽之比

$$Q\approx\frac{\omega_0\pi}{\mu A^2\Omega} \tag{5-4-26}$$

通常收敛因子μ取得很小，所以自适应陷波器的零点很深，带宽很窄，或者说Q值很高，而且能跟踪正弦干扰频率的变化，这是一般常规滤波器难以做到的。

§5-5 自适应波束形成器

先扼要地介绍一下常规波束形成器，看看常规波束形成器是如何提高输出信干比的。图5-5-1是常规波束形成器的原理图。

设有m个阵元组成的线列阵，有用信号源位于θ方向，干扰位于ϕ方向，由于干扰与信号具有相同的频谱，应用通常的频域滤波器是不可能消除干扰的，波束形成器可以看成是一个空间滤波器，利用干扰与信号在空间方向上的不同而滤除干扰。

设阵元本身是无方向性的，各阵元接收信号$s_1(t),s_2(t),\cdots,s_{m-1}(t),s_m(t)$，通过相应的延迟线，延迟时间分别为$\tau_1,$

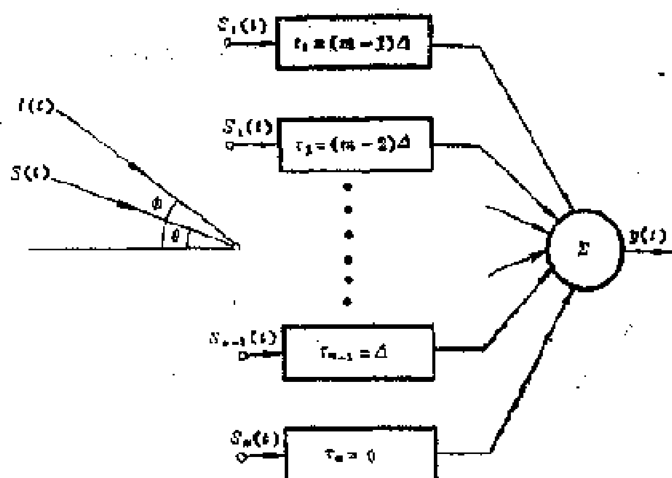


图5-5-1　常规波束形成器

$\tau_2,\cdots,\tau_{m-1},\tau_m$，令$\tau_m=(m-1)\Delta$，然后相加并输出。若选择$\Delta=\frac{d\sin\theta}{c}$，其中$c$为传播

速度，那么对于来自θ方向的有用信号，各阵元的接收信号将成为同相相加，而对于来自其它方向的干扰，则不构成同相相加，因此，输出端上的信干比增加，对于背景噪声同样也起到了信噪比提高的作用。

如果干扰很强，用常规波束形成得不到足够的输出信干比，这时就需要采用自适应波束形成器的方法。

图5-5-2是一种自适应波束形成器的原理图，它是由常规的波束形成器加一自适应干扰抵消器组成，抵消器的期望信号$d(t)$即为常规波束形成器的输出，抵消器的参考输入取自中间阵元的输出$s_{\frac{m+1}{2}}(t)$。

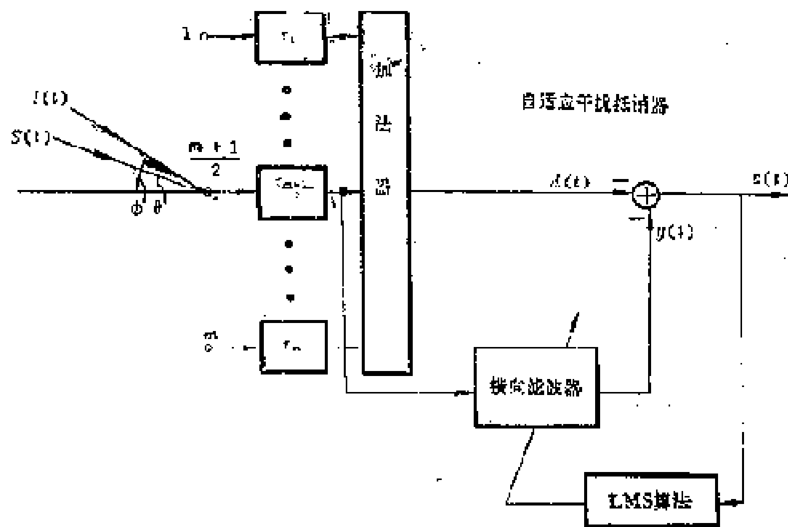


图5-5-2 无约束自适应波束形成器

因为单个阵元是无方向性的，所以参考信号的信干比很低，但是在输入$d(t)$中，经过常规波束形成器的处理，其信干比是相当高的，大约提高m倍。

根据前面对自适应干扰抵消器的讨论知道，在参考输入中不仅有干扰成分，而且也有有用信号的成分，自适应运算的结果，使$d(t)$中的干扰成分被大部分抵消，信号也有部分抵消，但是总的效果将使输出$\varepsilon(t)$中的信干比高于输入$d(t)$中的信干比。

图5-5-2中的自适应波束形成器结构具有简单和易于实现的优点，其缺点是有用信号被部分抑制，没有能够充分利用$d(t)$中的有用信号。

下面介绍另一种自适应波束形成器，称为带有引导信号的自适应波束形成器，图5-5-3是它的结构原理图，图5-5-4是其中处理器部分的详细原理图。

处理器由m个横向滤波器组成，其中的权都是可调的，并受自适应算法的控制。每个横向滤波器接到相应的阵元，对阵元的接收信号产生必要的相移或时延。

带有引导信号的自适应波束形成器的工作原理如下：它有一个引导信号发生器产生引导信号$d(t)$，一方面作为抵消器1的期望信号，一方面通过延迟线τ_1，τ_2，…，τ_m加到输入端，与对应阵元的接收信号叠加，在自适应运算过程中使均方误差最小化，也就是使$y(t)$去追近$d(t)$，同时尽可能地抑制与引导信号不相关的由阵元接收到的信号。因此，自适应波束形成器强迫方向性图在规定的引导方向上（由延迟线τ_1，τ_2，…，τ_m决定），在规定的带宽上（由引导信号频带决定）形成主波束。同时强迫阵方向性图在干扰

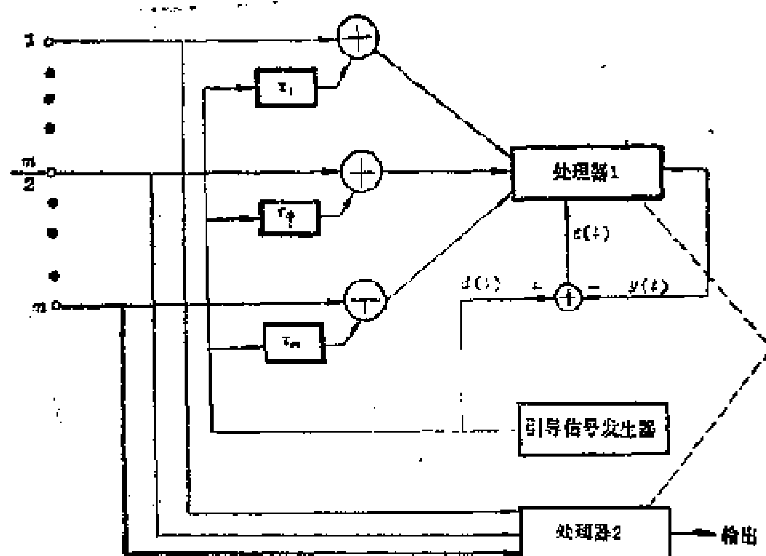


图5-5-3　带有引导信号的自适应波束形成器

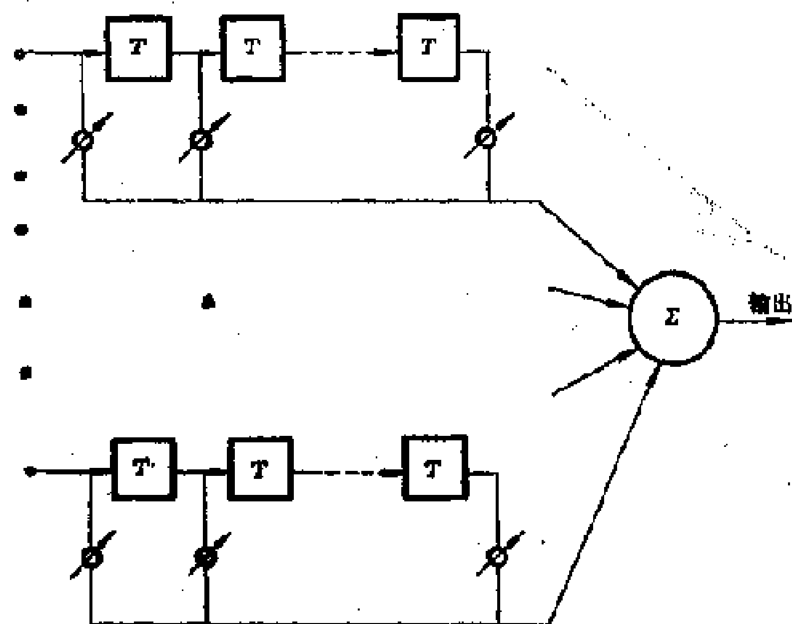


图5-5-4　宽带处理器

方向和其它频率上为零。由τ_1，τ_2，…，τ_m决定的方向称为束控方向。

在上述两种自应应波束形成器中，都假定干扰源是宽带信号，若干扰原是窄带信号，则抵消器的结构可以简化，在图5-5-4中的横向滤波器可以代之以一个复数权、m个横向滤波器代之以m个复数权。窄带干扰抵消器的结构如图5-5-5所示。

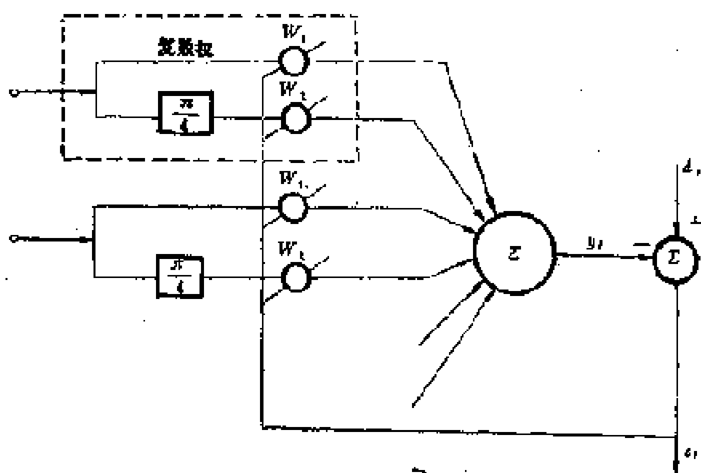
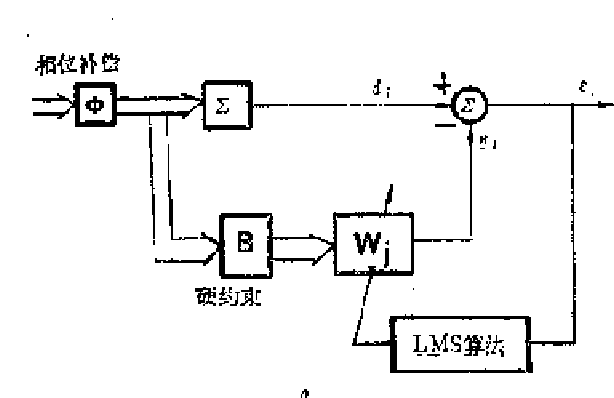


图5-5-5　窄带干扰抵消器

约束自适应波束形成器

上面讨论的自适应波束形成器称为无约束自适应波束形成器。下面介绍一种相减式硬约束的自适应基阵处理器，参看图5-5-6。

假设信号 $s(t)$与干扰 $I(t)$均为窄带信号；Φ 为相位补偿器，它确定常规阵束控方向，常规阵的输出作为抵消器的期望信号d_j，希望y_j中不含有 $s(t)$的成分，因此在横向滤波器之前加上约束矩阵 B。令



a

$$B=\begin{pmatrix} 1 & -1 & & & 0 \\ & 1 & 1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \\ 0 & & & 1 & -1 \end{pmatrix}$$

显然，对于有用信号 $s(t)$，相邻两路接收信号是同相位的，因而互相抵消。而对于另一方向来的干扰 $I(t)$，由于不同相位而不能抵消，因此横向滤波器的输入端只有干扰成分而没有信号成分，通过横向滤波器后成为y_j，y_j与d_j中的干扰相抵消，最后在输出ε_j中只包含有用信号成分。相减式硬约束自适应基阵处理器的结构如图5-5-6所示。

b

图5-5-6 相减式硬约束自适应基阵处理器

如果信号与干扰都是窄带的，则图中每个W代表一个复数权，如果信号与干扰都是宽带的，则图中每个W代表一个横向滤波器。

再介绍一种软约束自适应基阵处理器的原理。

在图5-5-7中，m路阵元信号通过延迟网络形成束控方向，这m路阵元信号经过相加器作为抵消器的期望信号 $d(t)$，其中有用信号增强了m倍，$d(t)$即为常规波束形成器的输出，这m路信号加到处理器，其结构仍如图5-5-4所示，由m个可变权的横向滤波器组成。我们希望 $y(t)$中不应含有用信号成分而只有干扰成分。那么，通过自适应运算，$d(t)$中的有用信号将不被抵消，而干扰得到很好的抵消，这时将得到很高的输出信干比。要使 $y(t)$不含有有用信号成分，就要使处理器中的权受到某种约束。下面讨论权的约束方法。

对照图5-5-4，处理器中的权按下列模式排列：

w_{11} w_{21} $\cdots$ w_{n1} ——横向滤波器 1

w_{12} w_{22} $\cdots$ w_{n2} ——横向滤波器 2

……………………………

w_{1m} w_{2m} $\cdots$ w_{nm} ——横向滤波器 m

若令

$$\left.\begin{aligned} w_{11}+w_{12}+\cdots+w_{1m}&=0\\ w_{21}+w_{22}+\cdots+w_{2m}&=0\\ &\cdots\cdots\\ w_{n1}+w_{n2}+\cdots+w_{nm}&=0 \end{aligned}\right\} \tag{5-5-1}$$

那末在$y(t)$中就没有束控方向的有用信号，这是因为束控方向的信号已被调整到同相，各路信号通过处理器时也是同步的，在上述约束条件下，就不能到达输出端。但对于干扰信号，由于空间方向不同，各路干扰通过处理器时是不同步的，因而能够到达输出端，所以，最终在$y(t)$中只有干扰信号而没有有用信号，因此$d(t)$中的有用信号不会被减去。

这种方法的主要困难在于需要采用约束自适应算法，其中权系数的迭代规则要比威特罗算法中的规则复杂得多。

关于自适应波束形成器就讲这几种，对于自适应波束形成器的研究，近年来发展甚快，还有很多种方法，但其基本工作原理都是以自适应干扰抵消器为基础的。大体上，要得到好的抵消效果常常要复杂的结构和算法，故在工程上大都采用次最佳实现方法。

自适应波束形成举例

下面给出一个自适应波束形成的模拟计算结果之例。图5-5-7为一12元水听器构成的圆阵及窄带自适应处理器单元，信号为单位幅度正弦波（频率f_0,功率0.5）；信号对准$\theta=0°$方向；干扰也为正弦波（频率f_0，功率0.5），干扰位于40°方向上，另外每路加有不相关的零均值白噪声，其方差等于0.1。

图5-5-7　圆阵及窄带适应自处理器单元

图5-5-8示出了自适应过程中方向性图的变化，每个方向性图是前各个时刻得到的一组权系数计算出来的。可以看到在$T=0$的初始时刻，方向性图是圆的，这是因为水听器阵元和加权是对

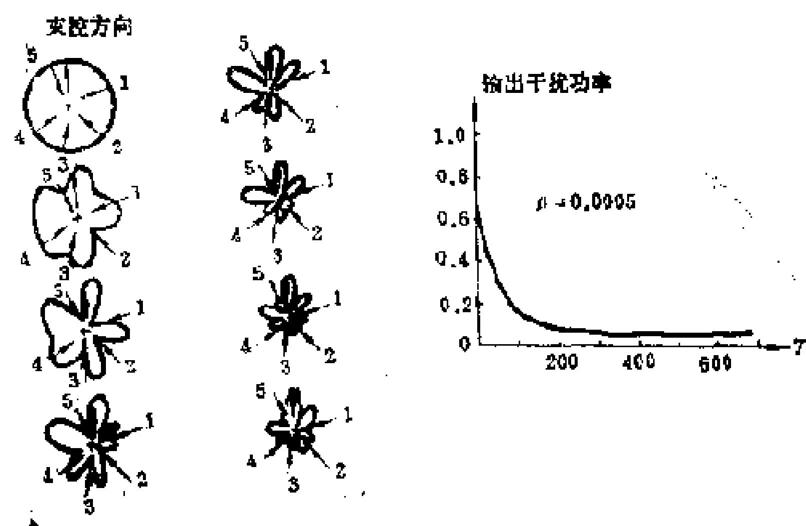
图5-5-8　自适应波束形成过程与学习曲线

称的，图中$T=\frac{1}{f_0}$为自适应过程时间。每个T期间进行20步迭代计算。

从图5-5-8中一系列方向性图可以得出一些看法，在束控方向上基阵灵敏度基本不变，而在干扰方向上基阵灵敏度迅速下降并形成一个陷谷，在自适应到达稳定后，干扰源方向的基阵灵敏度比束控方向的灵敏度大约低27dB。基阵输出的总干扰功率包括方向干扰正弦波与不相关的白噪声。总干扰功率随自适应过程的进行而下降，直到到达一个最低水平，图5-5-8还给出了总干扰功率与自适应时间的关系。

图5-5-9示出了有五个方向的正弦干扰，每个干扰的频率不相同，功率为0.125，另外加上不相关白噪声，每路功率为0.5。基阵与信号不变，图中示出了在自适应过程中方向性图的变化以及学习曲线。

由于各干扰的方向与频率是不相同的，故最终各干扰方向上基阵灵敏度也不相同，结果如下表所示。



干扰方向（度）	干扰频率（f_0的倍数）	干扰方向基阵灵敏度相对于束控方向(dB)
67°	1.10	−26
134°	0.95	−30
191°	1.00	−28
236°	0.90	−30
338°	1.05	28

图5-5-9　五个干扰源时自适应波束形成与学习曲线

图5-5-9中的方向性图是以中心频率为基础，根据各时刻上处理器中的权系数进行计算得到的。

上述例子告诉我们，自适应波束形成技术对于抑制不同方向上的同频干扰的能力是很强的，所以自适应波束形成技术受到人们极大的重视。

第三篇　卡尔曼滤波

第六章　恒增益递归滤波器与信号模型

对于离散时间随机过程进行滤波，滤波器的结构分两种类型：第一种类型是横向滤波器形式，也称非递归型滤波器，其特点是同时处理一批数据，当横向滤波器的权调节到最佳时，即成为维纳滤波器。第二种类型是递归型滤波器，其特点是每次只处理一个数据（标量或矢量），若递归滤波器中的反馈系数（又称为增益）取为某常值，则称为恒增益递归滤波器；若反馈系数根据信号模型中参数的变化随时调节到最佳值，即构成了卡尔曼滤波器。

本章中将首先研究恒增益递归滤波器，继而研究信号模型的性质，从下章起再研究卡尔曼滤波器。

§6-1　标量恒增益递归滤波器

从最简单的情况开始，设测量y_k是一标量平稳随机序列

$$y_k = x_k + e_k$$

其中x_k为信号，e_k为测量噪声，它们都是平稳随机过程，希望对x_k的均值进行估计，本章中用y_k而不用z_k表示测量序列，是为了避免与Z变换符号相混淆。

把y_k送入图6-1-1所示之一阶递归型滤波器，其输出作为估计$\hat{x}_k$，图a与图b是等价的。

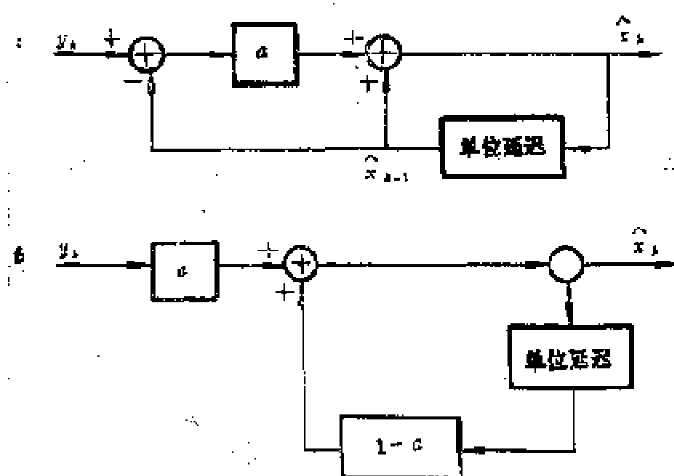


图6-1-1　一阶递归滤波器

由图6-1-1，我们得出

$$\hat{x}_k = \hat{x}_{k-1} + a(y_k - \hat{x}_{k-1}) \tag{6-1-1}$$

或者其等价形式

$$\hat{x}_k = (1-a)\hat{x}_{k-1} + a y_k \tag{6-1-2}$$

其中a是设定的常数，故滤波器是时不变的，a称为滤波器的增益。显见，在k时刻的滤波器输

出$\hat{x}_k$是前一时刻的$\hat{x}_{k-1}$加上一个修正量$a(y_k-\hat{x}_{k-1})$，我们把$(y_k-\hat{x}_{k-1})$称作k时刻的新息，意指测量值y_k中带来的新信息。由(6-1-1)式可见，滤波器具有反馈校正的作用，如差值$(y_k-\hat{x}_{k-1})$较大，则对$\hat{x}_{k-1}$的修正量也较大，反之，对$\hat{x}_{k-1}$的修正量变小。另外，由(6-1-2)式我们还可看到，$\hat{x}_k$是$\hat{x}_{k-1}$与y_k的线性组合。

递归滤波器存在稳定与否的问题，因此需要检查一下图6-1-1中滤波器的稳定性。设在输入端加上单位冲击$\delta(t)$，输出序列将是$a, a(1-a), a(1-a)^2, a(1-a)^3, \cdots$，显见，为使滤波器稳定，应选择$0<a<1$。

假定在$k=0$时刻之前，输入等于零，在$k=0$时刻之后，加上平稳序列$y_1, y_2, \cdots$，在各时刻滤波器的输出是

$$\hat{x}_1=(1-a)\hat{x}_0+ay_1$$

$$\hat{x}_2=(1-a)\hat{x}_1+ay_2$$

$$=(1-a)^2\hat{x}_0+a(1-a)y_1+ay_2$$

$$\cdots\cdots$$

$$\hat{x}_k=(1-a)\hat{x}_{k-1}+ay_k$$

$$=(1-a)^k\hat{x}_0+a(1-a)^{k-1}y_1+\cdots$$

$$+a(1-a)y_{k-1}+ay_k \qquad (6\text{-}1\text{-}3)$$

由于$0<a<1$，随着时间k的增长，初值$\hat{x}_0$和老的测量值对$\hat{x}_k$的影响越来越小，就好象滤波器对测量值有衰减记忆的功能。

考察$\hat{x}_k$的均值，对(6-1-3)式取统计平均，得

$$E[\hat{x}_k]=E[(1-a)^k\hat{x}_0+a(1-a)^{k-1}y_1+\cdots+a(1-a)y_{k-1}+ay_k]$$

$$=[(1-a)^k+a(1-a)^{k-1}+\cdots+a(1-a)+a]\overline{x}$$

$$=[(1-a)^k+a\{(1-a)^{k-1}+(1-a)^{k-2}+\cdots+(1-a)+1\}]\overline{x}$$

$$=\left[(1-a)^k+a\frac{1-(1-a)^k}{1-(1-a)}\right]\overline{x} \qquad (6\text{-}1\text{-}4)$$

当$k\to\infty$时，$E[\hat{x}_k]=\overline{x}$，由此可以得出结论，$\hat{x}_k$是渐近无偏估计。

考察$\hat{x}_k$的方差

$$P_k=E[(\hat{x}_k-\overline{x})^2]$$

$$=E[\{(1-a)(\hat{x}_{k-1}-\overline{x})+a(y_k-\overline{x})\}^2]$$

$$=(1-a)^2P_{k-1}+a^2\sigma_e^2 \qquad (6\text{-}1\text{-}5)$$

其中σ_e^2为测量噪声的方差，若给出初始估计方差P_0，则有

$$P_k=(1-a)^2P_{k-1}+a^2\sigma_e^2$$

$$=(1-a)^2[(1-a)^2P_{k-2}+a^2\sigma_e^2]+a^2\sigma_e^2$$

$$=(1-a)^4P_{k-2}+a^2(1-a)^2\sigma_e^2+a^2\sigma_e^2$$

$$\cdots\cdots$$

$$=(1-a)^{2k}P_0+a^2\sigma_e^2[(1-a)^{2(k-1)}+(1-a)^{2(k-2)}+\cdots$$

$$+(1-a)^2+1]$$

$$=(1-a)^{2k}P_0+a^2\sigma_v^2\frac{1-(1-a)^{2k}}{1-(1-a)^2} \tag{6-1-6}$$

令$k\to\infty$，取极限，得

$$\lim_{k\to\infty}P_k=\frac{a}{2-a}\sigma_v^2 \tag{6-1-7}$$

所以，当k足够大时，初值P_0的影响逐渐消失，即使P_0给得不准，也不会影响最后估计的质量。

再来考察滤波器的传递函数和频率响应，对（6-1-2）式取Z变换，得

$$x(Z)=(1-a)Z^{-1}\cdot x(Z)+ay(Z) \tag{6-1-8}$$

故滤波器的传递函数为

$$H(Z)=\frac{x(Z)}{y(Z)}=\frac{a}{1-(1-a)Z^{-1}} \tag{6-1-9}$$

再令，$Z=e^{j\omega T}$，就得出频率响应

$$H(\omega)=\frac{a}{1-(1-a)e^{-j\omega T}} \tag{6-1-10}$$

$$|H(\omega)|=\frac{a}{\{[1-(1-a)\cos\omega T]^2+(1-a)^2\sin^2\omega T\}^{1/2}}$$

$$=\frac{a}{[1+(1-a)^2-2(1-a)\cos\omega T]^{1/2}} \tag{6-1-11}$$

（6-1-11）式表明，这是一个低通滤波器，改变a值，可以改变滤波器的带宽。

X_k为非平稳情况

若信号x_k为非平稳随机过程，那末，

$$E[\hat{x}_k]=(1-a)E[\hat{x}_{k-1}]+aE[y_k]$$

$$=(1-a)\overline{x}_{k-1}+a\overline{x}_k \tag{6-1-12}$$

如果不知道$\overline{x}_{k-1}$和$\overline{x}_k$之间的关系，我们就无法计算$E[\hat{x}_k]$，也无法得到估计$\hat{x}_k$。

现在假定有下列关系式

$$x_k=\Phi x_{k-1}+w_{k-1} \tag{6-1-13}$$

其中Φ为常值，w_{k-1}为零均值白噪声。(6-1-13)式描述了信号状态变化的规律，称之为状态方程。因此，描述信号的完整模型由状态方程和测量方程组成

$$x_k=\Phi x_{k-1}+w_{k-1} \tag{6-1-14}$$

$$y_k=x_k+v_k \tag{6-1-15}$$

根据以上二式，我们可以把测量y_k看成是由白噪声w_{k-1}驱动一个线性滤波器所产生，如图6-1-2所示，称之为单输入单输出信号模型。

再把y_k送入图6-1-3所示之递归滤波器，就得到估计$\hat{x}_k$。

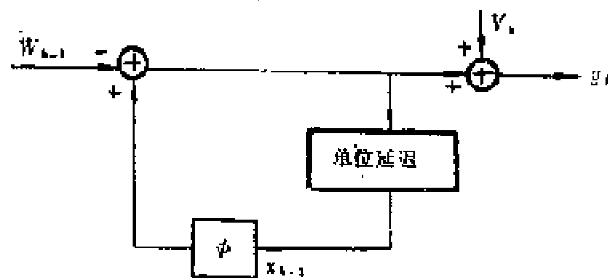


图6-1-2 单输入单输出信号模型

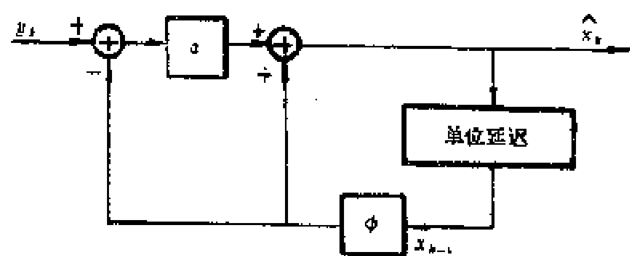


图6-1-3 单输入单输出递归滤波器

根据图6-1-3，得出

$$\hat{x}_k = \Phi\hat{x}_{k-1} + a[y_k - \Phi\hat{x}_{k-1}] \tag{6-1-16}$$

或

$$\begin{aligned}\hat{x}_k &= (1-a)\Phi\hat{x}_{k-1} + a y_k \\ &= (1-a)^k\Phi^k\hat{x}_0 \\ &\quad + (1-a)^{k-1}\Phi^{k-1}a y_1 + \cdots \\ &\quad + a y_k\end{aligned} \tag{6-1-17}$$

(6-1-16)或(6-1-17)式称为标量恒增益滤波方程。

考察$\hat{x}_k$的性质，即考察其均值与均方误差。对(6-1-17)式取统计平均，得

$$\begin{aligned}E[\hat{x}_k] &= E[(1-a)\Phi\hat{x}_{k-1} + a(x_k + v_k)] \\ &= E[(1-a)\Phi\hat{x}_{k-1} + a\Phi x_{k-1} \\ &\quad + a w_{k-1} + a v_k] \\ &= (1-a)\Phi E\hat{x}_{k-1} \\ &\quad + a\Phi E x_{k-1}\end{aligned} \tag{6-1-18}$$

若已知$\hat{x}_{k-1}$是x_{k-1}的无偏估计，则有

$$E[\hat{x}_k] = \Phi E\hat{x}_{k-1} = E[x_k] = \overline{x_k} \tag{6-1-19}$$

故$\hat{x}_k$是x_k的无偏估计。依此类推，只要使初始估计$\hat{x}_0 = \overline{x_0}$，则每一步估计都是无偏的。倘若，$\hat{x}_0 \neq \overline{x_0}$，则估计不是无偏的。

相应的均方误差为

$$\begin{aligned}P_k &= E[(\hat{x}_k - x_k)^2] \\ &= E[\{(1-a)\Phi(\hat{x}_{k-1} - x_{k-1}) - (1-a)w_{k-1} + a v_{k-1}\}^2]\end{aligned}$$

因为估计误差$\hat{x}_{k-1} - x_{k-1}$，噪声w_{k-1}和v_k已假定彼此不相关，令$E[w_k^2] = Q_k$，$E[v_k^2] = R$则

$$P_k = (1-a)^2\Phi^2 P_{k-1} + (1-a)^2 Q_{k-1} + a^2 R_k \tag{6-1-20}$$

依此类推，可得知P_k是P_0，Q_k和R_k的线性组合。显见，只要满足条件

$$(1-a)\Phi < 1, \qquad 0 < a < 1 \tag{6-1-21}$$

则随着k的增长，初值P_0的影响会逐渐消失。此外，根据(6-1-7)式，容易推知，在(6-1-21)式条件下，初始估计$\hat{x}_0$的影响也会逐渐消失。

故图6-1-3中的递归滤波器成为稳定的条件是$0 < a < 1$和$(1-a)\Phi < 1$。

当x_k为非平稳过程时，可得出下述结论：

1． 当x_k为非平稳过程时，需要知道描述x_k变化规律的状态方程，才能构造相应的滤波器。

2. 常数a称为滤波器增益，a的取值决定着滤波器的带宽和估计的性能。

3. 滤波器成为稳定的充分的必要条件是$0<a<1$和$(1-a)\Phi<1$。对于稳定的滤波器，初值$\hat{x}_0$和P_0的影响随着时间的增长而逐渐消失。

4. 若给定初值$\hat{x}_0=\bar{x}_0$，则$\hat{x}_k$成为x_k的无偏估计。

§6-2 α-β 滤 波 器

现在把标量恒增益递归滤波器的结果推广到矢量情况。$\alpha-\beta$滤波器是一种常见的和用处很大的恒增益递归滤波器，下面就来讨论$\alpha-\beta$滤波器。

假设有一等速运动目标，但附加有随机速度扰动，目标在k时刻的运动状态可用距离变量x_k和速度变量$\dot{x}_k$表示之，根据牛顿力学有

$$x_{k+1}=x_k+T\dot{x}_k \tag{6-2-1}$$

$$\dot{x}_{k+1}=\dot{x}_k+w_k \tag{6-2-2}$$

其中w_k是随机速度扰动，通常把w_k看成是零均值白噪声。

定义状态矢量为$X=\begin{pmatrix}x_k\\ \dot{x}_k\end{pmatrix}$，把(6-2-1)和(6-2-2)式合并写成矢量方程

$$\begin{pmatrix}x_{k+1}\\ \dot{x}_{k+1}\end{pmatrix}=\begin{pmatrix}1 & T\\ 0 & 1\end{pmatrix}\begin{pmatrix}x_k\\ \dot{x}_k\end{pmatrix}+\begin{pmatrix}0\\ 1\end{pmatrix}w_k \tag{6-2-3}$$

记 $\Phi=\begin{pmatrix}1 & T\\ 0 & 1\end{pmatrix}$，$\Phi$称为状态转移矩阵。

$\Gamma=\begin{pmatrix}0\\ 1\end{pmatrix}$，$\Gamma$称为状态噪声转移矩阵。

把(6-2-3)式重写成

$$X_{k+1}=\Phi X_k+\Gamma w_k \tag{6-2-4}$$

(6-2-4)式即为描述目标状态变化规律的状态方程。

设测量值只是距离，测量方程为

$$y_k=x_k+v_k \tag{6-2-5}$$

其中v_k为测量噪声，假定它也是零均值白噪声，并且与w_k不相关，为了把测量值与状态矢量联系起来，把(6-2-5)式写成

$$y_k=[1,\ 0]\begin{pmatrix}x_k\\ \dot{x}_k\end{pmatrix}+v_k \tag{6-2-6}$$

记$[1,\ 0]=H$，H称为测量矩阵，则有

$$y_k = H X_k + v_k \tag{6-2-7}$$

(6-2-4) 与 (6-2-7) 式即构成了信号模型，现在让测量y_k通过一恒增益递归滤波器，以便得出状态矢量 X_k在各时刻的估计。我们现在来构造这个滤波器。

如果已经知道了k时刻目标状态的估计为$\hat{X}_k$，那末由状态方程 (6-2-3) 式可以预报$k+1$时刻的状态估计，记预报估计为$\hat{X}_{k+1/k}$

$$\hat{X}_{k+1/k} = \Phi \hat{X}_k = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \hat{X}_k \tag{6-2-8}$$

(6-2-8)式称为预报方程。

在得到新测量值y_k之后，需要对预报估计进行修正，以便得到较好的估计。由y_k带来的距离新息是$(y_{k+1} - \hat{x}_{k+1/k})$，由$y_k$带来的速度新息是 $\frac{y_{k+1} - \hat{x}_{k+1/k}}{T}$。参照上节中的标量递归滤波器方程，得出$k+1$时刻的估计为

$$\hat{x}_{k+1} = \hat{x}_{k+1/k} + \alpha(y_{k+1} - \hat{x}_{k+1/k}) \tag{6-2-9}$$

$$\hat{\dot{x}}_{k+1} = \hat{\dot{x}}_{k+1/k} + \beta\left(\frac{y_{k+1} - \hat{x}_{k+1/k}}{T}\right) \tag{6-2-10}$$

合并成矢量形式

$$\hat{X}_{k+1} = \hat{X}_{k+1/k} + K[y_{k+1} - H\hat{X}_{k+1/k}] \tag{6-2-11}$$

其中 $K = \begin{pmatrix} \alpha \\ \frac{\beta}{T} \end{pmatrix}$，$K$称为滤波器的增益矢量，(6-2-11)式称为滤波方程。

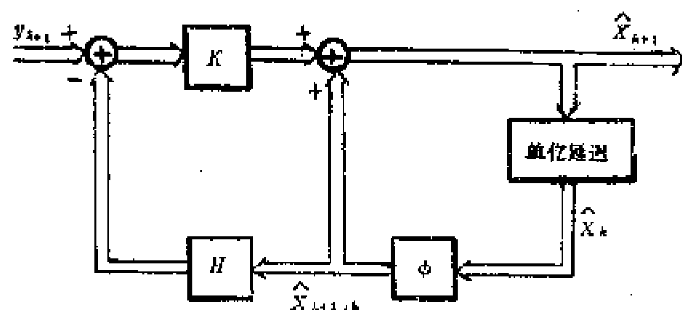


图6-2-1　矢量递归滤波器

根据 (6-2-8) 和(6-2-11)式，可以构造出矢量恒增益递归滤波器，如图 6-2-1 所示，这是单输入双输出的递归型滤波器。图6-2-1中的滤波器或由(6-2-8)和(6-2-11)式表达的滤波器在工程上常常称为$\alpha-\beta$滤波器，下面我们来研究$\alpha-\beta$滤波器的各项性能。

1. α-β滤波器的传递函数与频率响应

把滤波方程改写成

$$\begin{aligned}\hat{X}_k &= (1-KH)\Phi\hat{X}_{k-1} + Ky_k \\ &= A\hat{X}_{k-1} + Ky_k\end{aligned} \tag{6-2-12}$$

其中 $A=(1-KH)\Phi=\begin{pmatrix}1-\alpha & 0\\ -\dfrac{\beta}{T} & 1\end{pmatrix}\begin{pmatrix}1 & T\\ 0 & 1\end{pmatrix}$

$$=\begin{pmatrix}1-\alpha & (1-\alpha)T\\ -\dfrac{\beta}{T} & 1-\beta\end{pmatrix} \tag{6-2-13}$$

$$K=\begin{pmatrix}\alpha\\ \beta/T\end{pmatrix} \tag{6-2-14}$$

对(6-2-12)式两边进行Z变换，得

$$\hat{X}(Z)=AZ^{-1}\hat{X}(Z)+Ky(Z) \tag{6-2-15}$$

故$\alpha-\beta$滤波器的传递函数为

$$H(Z)=\frac{\hat{X}(Z)}{y(Z)}=(1-AZ^{-1})^{-1}K \tag{6-2-16}$$

其中 $1-AZ^{-1}=\begin{pmatrix}1-(1-\alpha)Z^{-1} & -(1-\alpha)TZ^{-1}\\ \dfrac{\beta}{T}Z^{-1} & 1-(1-\beta)Z^{-1}\end{pmatrix}$

$$(1-AZ^{-1})^{-1}=\frac{\begin{pmatrix}1-(1-\beta)Z^{-1} & (1-\alpha)TZ^{-1}\\ -\dfrac{\beta}{T}Z^{-1} & 1-(1-\alpha)Z^{-1}\end{pmatrix}}{1-(2-\alpha-\beta)Z^{-1}+(1-\alpha)Z^{-2}} \tag{6-2-17}$$

因为 $$H(Z)=\frac{\hat{X}(Z)}{y(Z)}=\begin{pmatrix}\dfrac{\hat{x}(Z)}{y(Z)}\\ \dfrac{\hat{\dot{x}}(Z)}{y(Z)}\end{pmatrix}=\begin{pmatrix}H_{\hat{x}}(Z)\\ H_{\hat{\dot{x}}}(Z)\end{pmatrix} \tag{6-2-18}$$

我们可把$H(Z)$分解成两部分，一个是$H_{\hat{x}}(Z)$，另一个是$H_{\hat{\dot{x}}}(Z)$。

$$H_{\hat{x}}(Z)=\frac{\hat{x}(Z)}{y(Z)}=\frac{\alpha+(\beta-\alpha)Z^{-1}}{1-(2-\alpha-\beta)Z^{-1}+(1-\alpha)Z^{-2}}$$

$$=\frac{\alpha Z\left(Z+\dfrac{\beta-\alpha}{2}\right)}{Z^2-(2-\alpha-\beta)Z+(1-\alpha)} \tag{6-2-19}$$

$$H_{\hat{\dot{x}}}(Z)=\frac{\hat{\dot{x}}(Z)}{y(Z)}=\frac{\dfrac{\beta}{T}(1-Z^{-1})}{1-(2-\alpha-\beta)Z^{-1}+(1-\alpha)Z^{-2}}$$

$$= \frac{\frac{\beta}{T}Z(Z-1)}{Z^2-(2-\alpha-\beta)Z+(1-\alpha)} \tag{6-2-20}$$

此外，对于距离预报，也可以求出其传递函数

$$H_{\hat{x}_{k+1/k}}(Z) = \frac{\hat{x}(Z)+T\,\hat{\dot{x}}(Z)}{y(Z)}$$

$$= H_{\hat{x}}(Z) + T\,H_{\hat{\dot{x}}}(Z) \tag{6-2-21}$$

令 $Z=e^{j\omega T}$，代入 (6-2-19)和(6-2-20)两式中，就可得相应的滤波器频率响应$H_{\hat{x}}(\omega)$和$H_{\hat{\dot{x}}}(\omega)$，再取其模就得到幅频响应$|H_{\hat{x}}(\omega)|$和$|H_{\hat{\dot{x}}}(\omega)|$。图6-2-2中示出了$\alpha=0.9$，$\beta=1.0$，0.5，0.1和$\alpha=0.1$，$\beta=0.1$，0.05，0.01时两组$|H_{\hat{x}}(\omega)|$和$|H_{\hat{\dot{x}}}(\omega)|$的幅频响应。显见，$|H_{\hat{x}}(\omega)|$具有低通滤波器的性质，而$H_{\hat{\dot{x}}}(\omega)$具有带通滤波器性质。

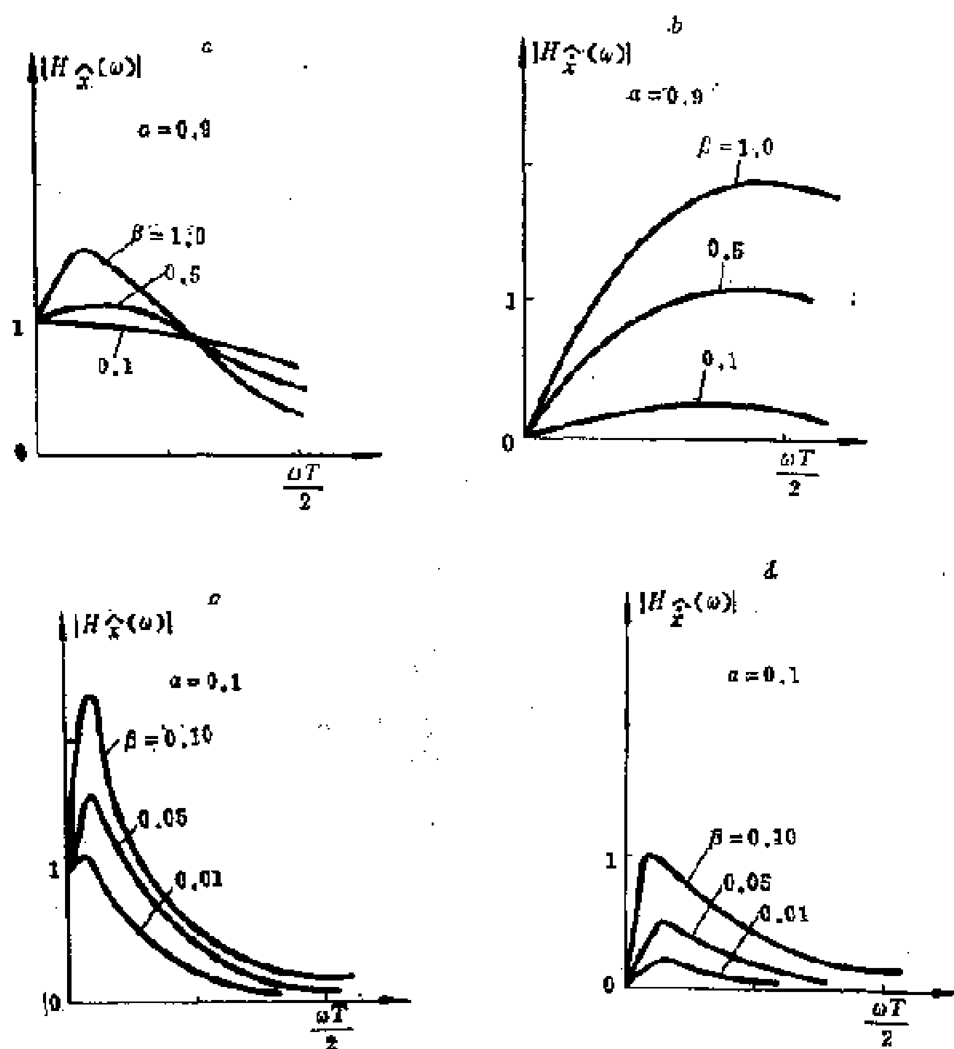


图6-2-2　α-β滤波器的幅频响应

2. α-β滤波器的硬件结构

图6-2-1画出了α-β滤波器的矢量结构形式，是难以具体实现的。根据$H_{\hat{x}}(Z)$，$H_{\hat{\dot{x}}}(Z)$，$H_{\hat{x}_{k+1/k}}(Z)$的表达式可以构造出α-β滤波器的硬件结构。分开构造三个滤波器太浪费，可以合并成一个，如图6-2-3所示。当数据采样率比较高，用软件进行滤波处理速度不够时，就可采用硬件处理。

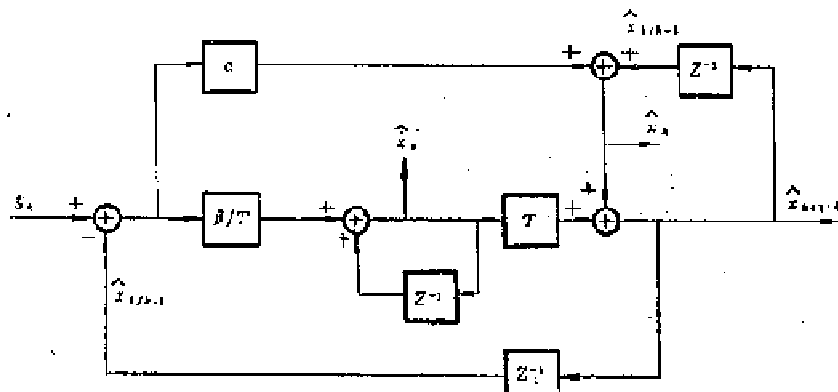


图6-2-3　α-β滤波器信号流图

3. α-β滤波器的稳定性与参数选择

因为α-β滤波器是一反馈系统，必然存在稳定性问题。为此必须正确选择参数α和β之值，以保证滤波器是稳定的。重写α-β滤波器表达式如下

$$\hat{x}_{k+1} = \hat{x}_{k+1/k} + \alpha(y_{k+1} - \hat{x}_{k+1/k}) \tag{6-2-22}$$

$$\hat{\dot{x}}_{k+1} = \hat{\dot{x}}_{k+1/k} + \beta\left(\frac{y_{k+1} - \hat{x}_{k+1/k}}{T}\right) \tag{6-2-23}$$

若输入y_{k+1}为单位冲击，为使（6-2-22）、（6-2-23）式稳定。根据第1节所述理由，应选择

$$0 < \alpha < 1, \qquad 0 < \beta < 1 \tag{6-2-24}$$

另外，从传递函数$H_{\hat{x}}(Z)$和$H_{\hat{\dot{x}}}(Z)$的表达式可以看到，它们的共同分母是一个特征方程

$$Z^2 - (2 - \alpha - \beta)Z + 1 - \alpha = 0 \tag{6-2-25}$$

特征方程的根乃是传递函数的极点

$$Z_{1,2} = \frac{(2-\alpha-\beta) \pm \sqrt{(2-\alpha-\beta)^2 - 4(1-\alpha)}}{2}$$

$$= \frac{(2-\alpha-\beta) \pm \sqrt{(\alpha+\beta)^2 - 4\beta}}{2} \tag{6-2-26}$$

只要Z_1、Z_2位于Z平面中单位圆内，滤波器就是稳定的。

由(6-2-16)式容易看出，极点Z_1、Z_2就是矩阵$A=(1-KH)\Phi$的特征值。

$$A = (1-KH)\Phi = \begin{pmatrix} 1-\alpha & (1-\alpha)T \\ -\dfrac{\beta}{T} & 1-\beta \end{pmatrix} \tag{6-2-27}$$

写出特征方程

$$|\lambda I - A| = \begin{vmatrix} \lambda - 1 + \alpha & -(1-\alpha)T \\ -\dfrac{\beta}{T} & \lambda - 1 + \beta \end{vmatrix} = 0 \tag{6-2-28}$$

或

$$(\lambda + \alpha - 1)(\lambda + \beta - 1) + \beta(1 - \alpha) = 0$$

$$\lambda^2 - (2 - \alpha - \beta)\lambda + 1 - \alpha = 0$$

解出A的特征值

$$\lambda_{1,2} = \frac{(2 - \alpha - \beta) \pm \sqrt{(2 - \alpha - \beta)^2 - 4(1 - \alpha)}}{2} \tag{6-2-29}$$

所以，滤波器稳定条件可表达为要求矩阵$A = (I - KH)\Phi$的特征值位于单位圆内。

根据极点Z_1、Z_2在单位圆内的分布，稳定性分三种类型：

(1) 若 $0 < \alpha < 1$，$0 < \beta < 1$，$(\alpha + \beta)^2 > 4\beta$ (6-2-30)

则Z_1、Z_2为正实极点，属过阻尼状态。

(2) 若 $0 < \alpha < 1$，$0 < \beta < 1$，$(\alpha + \beta)^2 < 4\beta$ (6-2-31)

此时，Z_1，Z_2为一对共轭极点，其实部为正，属欠阻尼状态。

(3) 若 $0 < \alpha < 1$，$0 < \beta < 1$，$(\alpha + \beta)^2 = 4\beta$ (6-2-32)

此时，$Z_1 = Z_2$，为双重正实极点，属临界阻尼状态。通常，可取临界阻尼状态作为选择滤波器的参数α、β的条件。因此，得出α、β必须满足的条件为

$$0 < \alpha < 1, \quad 0 < \beta < 1, \quad (\alpha + \beta)^2 = 4\beta \tag{6-2-33}$$

注意，这里只给出了α、β的取值范围，在这个范围内能保证滤波器是稳定的，但并不保证滤波器是最佳的，为了得到最佳滤波器增益，需要应用卡尔曼滤波理论，这个问题留在以后研究。

4. 估计的均值与均方误差

在滤波器稳定的条件下再研究估计的性能，即估计$\hat{X}_k$的均值和均方误差。

首先考察均值$E[\hat{X}_k]$。假定前一时刻的估计$\hat{X}_{k-1}$是无偏的，$E[\hat{X}_{k-1}] = \overline{X}_{k-1}$。因为

$$\begin{aligned} \hat{X}_k &= (I - KH)\Phi\hat{X}_{k-1} + Ky_k \\ &= (I - KH)\Phi\hat{X}_{k-1} + KH(\Phi X_{k-1} + \Gamma w_{k-1}) + Kv_k \end{aligned} \tag{6-2-34}$$

两边取统计平均，得

$$E[\hat{X}_k] = \Phi E[\hat{X}_{k-1}] = \Phi \overline{X}_{k-1} = \overline{X}_k \tag{6-2-35}$$

所以，$\hat{X}_k$为X_k的无偏估计。只要初值$\hat{X}_0 = \overline{X}_0$，则每一步估计都是无偏估计，这里$\overline{X}_0$为初始状态$X_0$的均值。

再考察估计的均方误差。定义估计误差为

$$\begin{aligned} \tilde{X}_k &= X_k - \hat{X}_k \\ &= (I - KH)\tilde{X}_{k/k-1} - Kv_k \end{aligned} \tag{6-2-36}$$

由此得到估计的均方误差为

$$P_k = E[\widetilde{X}_k \widetilde{X}_k^T] \tag{6-2-37}$$

考察到在$\widetilde{X}_{k-1}$、w_{k-1}、v_k之间相互独立的条件下，得均方误差为

$$P_k = (1-KH)(\Phi P_{k-1}\Phi^T + \Gamma Q_{k-1}\Gamma^T)(1-KH)^T + KR_kK^T \tag{6-2-38}$$

其中 $Q_{k-1} = E[w_{k-1}w_{k-1}^T]$， $R_k = E[v_kv_k^T]$，

$$K = \begin{pmatrix} \alpha \\ \dfrac{\beta}{T} \end{pmatrix}$$

迄今得到了两个重要的递推公式

$$\hat{X}_k = (1-KH)\Phi\hat{X}_{k-1} + Ky_k \tag{6-2-39}$$

$$P_k = (1-KH)(\Phi P_{k-1}\Phi^T + \Gamma Q_{k-1}\Gamma^T)(1-KH)^T + KR_kK^T \tag{6-2-40}$$

在选定滤波器参数α、β之值后，给出初始条件$\hat{X}_0$和P_0，根据各个时刻的测量值y_k，测量噪声方差R_k和状态噪声方差Q_k，就可递推地计算出各个时刻的估计$\hat{X}_k$及其均方误差P_k。

§6-3 α-β-γ 滤波器

若目标作等加速运动，但受到随机加速度扰动的作用，由质点动力学可得出下列关系

$$x_{k+1} = x_k + T\dot{x}_k + \frac{T^2}{2}\ddot{x}_k \tag{6-3-1}$$

$$\dot{x}_{k+1} = \dot{x}_k + T\ddot{x}_k \tag{6-3-2}$$

$$\ddot{x}_{k+1} = \ddot{x}_k + w_k \tag{6-3-3}$$

合并成矢量形式，得

$$\begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \end{pmatrix}_{k+1} = \begin{pmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \end{pmatrix}_k + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} w_k \tag{6-3-4}$$

记

$$\Phi = \begin{pmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & 1 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad X_k = \begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \end{pmatrix}_k$$

则有

$$X_{k+1} = \Phi X_k + \Gamma w_k \tag{6-3-5}$$

（6-3-5）式就是该目标的状态方程。

假定测量值是距离，则可得出测量方程为

$$y_k=[1,\ 0,\ 0]\begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \end{pmatrix}_k+v_k \tag{6-3-6}$$

记 $H=[1,\ 0,\ 0]$，则得出测量方程为

$$y_k=HX_k+v_k \tag{6-3-7}$$

(6-3-5)和(6-3-6)式构成了信号模型。

现在构造一个三维矢量滤波器，输入是测量值y_k，输出应当是x_k、$\dot{x}_k$、$\ddot{x}_k$的合理估计$\hat{x}_k$、$\hat{\dot{x}}_k$和$\hat{\ddot{x}}_k$。

假定已得到k时刻的状态矢量估计$\hat{X}_k$，那末由状态方程可以得出$k+1$时刻的预报估计$\hat{X}_{k+1/k}=\Phi\hat{X}_k$，而由测量值$y_{k+1}$带来的距离新息是$(y_{k+1}-\hat{x}_{k+1/k})$，速度新息是$\frac{1}{T}(y_{k+1}-\hat{x}_{k+1/k})$，加速度新息是$\frac{2}{T^2}(y_{k+1}-\hat{x}_{k+1/k})$。于是构造滤波器方程如下

$$\hat{x}_{k+1}=\hat{x}_{k+1/k}+\alpha(y_{k+1}-\hat{x}_{k+1/k}) \tag{6-3-8}$$

$$\hat{\dot{x}}_{k+1}=\hat{\dot{x}}_{k+1/k}+\beta\left(\frac{y_{k+1}-\hat{x}_{k+1/k}}{T}\right) \tag{6-3-9}$$

$$\hat{\ddot{x}}_{k+1}=\hat{\ddot{x}}_{k+1/k}+\gamma\left(\frac{2}{T^2}(y_{k+1}-\hat{x}_{k+1/k})\right) \tag{6-3-10}$$

写成矢量形式为

$$\hat{X}_{k+1}=\hat{X}_{k+1/k}+K[y_{k+1}-H\hat{X}_{k+1/k}] \tag{6-3-11}$$

其中 $\hat{X}_{k+1/k}=\Phi\hat{X}_k \qquad K=\begin{bmatrix} \alpha \\ \beta/T \\ 2\gamma/T^2 \end{bmatrix}$

$H=[1,\ 0,\ 0]$

(6-3-11)式称为α-β-γ滤波器。下面研究α-β-γ滤波器的性能。

1. 传递函数与频率响应

采用相同于α-β滤波器中的方法，对滤波器方程（6-3-11）式进行Z变换，就得出α-β-γ滤波器的传递函数

$$H_{\hat{x}}(Z)=\frac{\alpha Z^3-(2\alpha-\beta-\gamma)Z+(\alpha-\beta+\gamma)}{Z^3-(3-\alpha-\beta-\gamma)Z^2+(3-2\alpha-\beta+\gamma)Z-(1-\alpha)} \tag{6-3-12}$$

$$H_{\hat{\dot{x}}}(Z)=\frac{\left(\frac{\beta}{T}Z\right)\left(Z-1\right)\left(Z+\frac{2\alpha-\beta}{T}\right)}{Z^3-(3-\alpha-\beta-\gamma)Z^2+(3-2\alpha-\beta+\gamma)Z-(1-\alpha)} \tag{6-3-13}$$

$$H_{\hat{\ddot{x}}}(Z)=\frac{\left(\frac{2\gamma}{T^2}Z\right)\left(Z-1\right)^2}{Z^3-(3-\alpha-\beta-\gamma)Z^2-(3-2\alpha-\beta+\gamma)Z-(1-\alpha)} \tag{6-3-14}$$

令$Z=e^{j\omega T}$，代入（6-2-12）、（6-2-13）、（6-2-14）式中，就得出滤波器的频率响应，再取模就得出其幅频响应$|H_{\hat{x}}(\omega)|$、$|H_{\hat{\dot{x}}}(\omega)|$和$|H_{\hat{\ddot{x}}}(\omega)|$。三种幅频响应的形状大体如图6-3-1所示，其中$|H_{\hat{x}}(\omega)|$为低通滤波器，$|H_{\hat{\dot{x}}}(\omega)|$和$|H_{\hat{\ddot{x}}}(\omega)|$为带通滤波器，$\omega_s$为采样角频率。

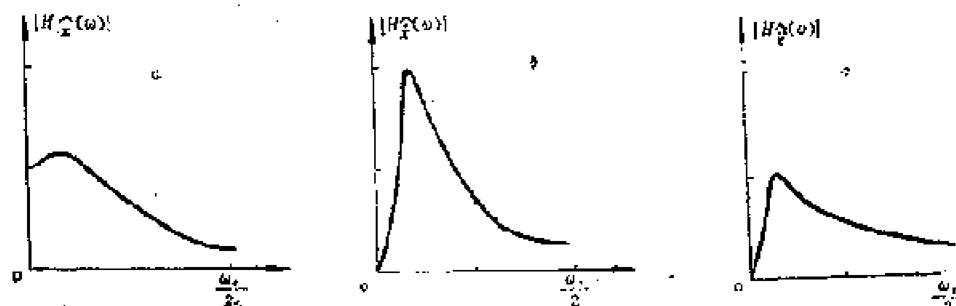


图6-3-1 滤波器的幅频响应

2. α-β-γ滤波器的稳定性与参数选择

已知各传递函数的特征方程为

$$D(Z)=Z^3-(3-\alpha-\beta-\gamma)Z^2+(3-2\alpha-\beta+\gamma)Z-(1-\alpha)=0 \qquad (6\text{-}3\text{-}15)$$

在临界阻尼时，特征方程的根是三重正实根，设此根的值为R，由代数学知，α、β、γ与R的关系为$D(Z)=(Z-R)^3=Z^3-3RZ^2+3R^2Z-R^3$。由此解得

$$\left.\begin{aligned}\alpha&=1-R^3\\ \beta&=1.5(1-R^2)(1-R)\\ \gamma&=0.5(1-R)^3\end{aligned}\right\} \qquad (6\text{-}3\text{-}16)$$

为了保证滤波器的稳定性，要求R位于Z平面的单位圆中，故应满足条件$R<1$。此外，由滤波器方程（6-3-8）、（6-3-9）、（6-3-10）式，根据与上节同样的理由，要求$0<\alpha<1$，$0<\beta<1$，$0<\gamma<1$，再选定R之值，就可按照（6-3-16）式确定相应的滤波器参数α、β、γ之值。

根据上述原则设计的α-β-γ滤波器，只保证滤波器是稳定的，但并不保证滤波器是最佳的。

由于α-β-γ滤波器与α-β滤波器的结构完全相似，只是维数不同。所以，α-β-γ滤波器的性能亦与α-β滤波器相似，不再赘述。α-β-γ滤波器中$\hat{X}_k$和P_k的递推公式亦仍为(6-2-39)和(6-2-40)式，只是Φ、Γ、H、K的表达式不同。

§6-4 信号状态变量模型

通过以上各节的讨论，可以看到，描述时变过程需要用两个方程式，一个是描述信号状态变化规律的状态方程，一个是把测量值与状态矢量联系起来的测量方程。对于线性时变过程，其状态方程和测量方程的一般形式为

$$X_{k+1}=\Phi_{k+1/k}X_k+B_kU_k+\Gamma_kW_k \qquad (6\text{-}4\text{-}1)$$

$$Z_{k+1}=H_{k+1}X_{k+1}+V_{k+1} \tag{6-4-2}$$

其中 $\Phi_{k+1/k}$为状态转移矩阵，

U_k为外加控制量，

B_k为控制转移矩阵，

W_k为状态噪声，

Γ_k为噪声转移矩阵，

H_{k+1}为测量矩阵，

V_{k+1}为测量噪声。

假定W_k、V_k均为零均值白噪声，且互不相关。

测量Z_{k+1}可以看成是由W_k和U_k通过一个线性系统所产生，可以用下面的框图表示之。

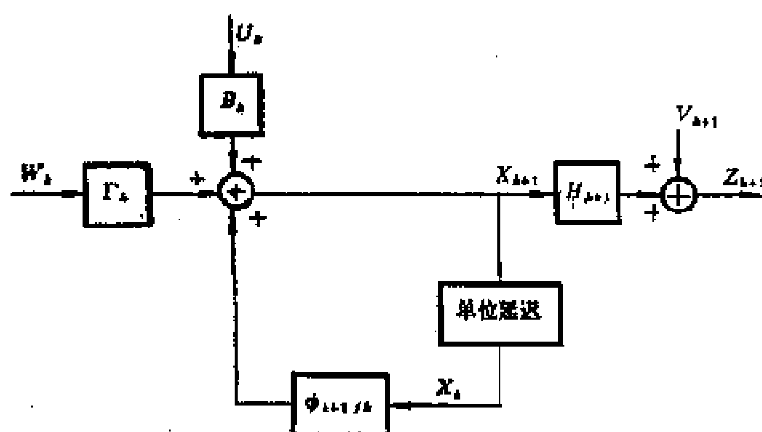


图6-4-1 矢量线性信号模型

图6-4-1或方程式(6-4-1)、(6-4-2)称为线性信号模型，线性信号模型分定常与非定常两大类。对于(6-4-1)、(6-4-2)式所示之模型，其中$\Phi_{k+1/k}$、B_k、Γ_k、H_{k+1}，状态噪声的方差$E[W_kW_k^T]=Q_k$和测量噪声方差$E[V_kV_k^T]=R_k$与时间k有关，则称为非定常模型，若参数 $\Phi_{k+1/k}=\Phi$，$B_k=B$，$\Gamma_k=\Gamma$，$H_k=H$，$Q_k=Q$，$R_k=R$，均与时间k无关，则此模型称为定常模型。在上述参数中只要有一个与时间k有关就是非定常模型。

若$\Phi_{k+1/k}$、B_k、Γ_k、H_k等参数中任一个为状态X_k的非线性函数，则称为非线性模型，非线性模型的表达式如下

$$X_{k+1}=\Phi(X_k)+B(X_k)U_k+\Gamma(X_k)W_k \tag{6-4-3}$$

$$Z_{k+1}=h(X_{k+1})-V_{k+1} \tag{6-4-4}$$

本章中将研究线性模型的各项重要性质。

状态方程的解

因为状态方程描述了状态的变化规律，故在k时刻的状态X_k可以用以前的输入W_i和初始状态X_0加以表示。

首先考察定常线性模型，设状态方程为

$$X_k=\Phi X_{k-1}+BU_{k-1}+\Gamma W_{k-1} \tag{6-4-5}$$

已知初始状态为X_0，在$k=1$，2，…各时刻，控制输入和噪声输入为(U_1, W_1)，(U_2, W_2)，…。利用递推法，可以求出各个时刻的状态矢量X_k

$k=1$ 时，　$X_1=\Phi X_0$，X_0为初值

$k=2$ 时，　$X_2=\Phi X_1+BU_1+\Gamma W_1$

$$=\Phi^2X_0+BU_1+\Gamma W_1$$

$k=3$ 时，　$X_3=\Phi X_2+BU_2+\Gamma W_2$

$$=\Phi^3X_0+\Phi BU_1+BU_2+\Phi\Gamma W_1+BW_2$$

…………………………

$$k=k\text{ 时，}\quad X_k=\Phi^kX_0+\sum_{i=0}^{k-1}\Phi^{k-i-1}BU_i+\sum_{i=0}^{k-1}\Phi^{k-i-1}\Gamma W_i \qquad (6\text{-}4\text{-}6)$$

(6-4-6)式称为状态方程的解。

对于非定常模型

$$X_k=\Phi_{k/k-1}X_{k-1}+B_{k-1}U_{k-1}+\Gamma_{k-1}W_{k-1} \qquad (6\text{-}4\text{-}7)$$

同理可以得出状态方程的解为

$$X_k=\Phi_{k/k-1}[\Phi_{k-1/k-2}X_{k-2}+B_{k-2}U_{k-2}+\Gamma_{k-2}W_{k-2}]$$
$$+B_{k-1}U_{k-1}+\Gamma_{k-1}W_{k-1}$$

…………………………

$$=\Phi_{k/k-1}\Phi_{k-1/k-2}\cdots\Phi_{1/0}X_0$$
$$+\sum_{i=0}^{k-2}\Phi_{k/k-1}\Phi_{k-1/k-2}\cdots\Phi_{i+2/i+1}B_iU_i+B_{k-1}U_{k-1}$$
$$+\sum_{i=0}^{k-2}\Phi_{k/k-1}\Phi_{k-1/k-2}\cdots\Phi_{i+2/i+1}\Gamma_iW_i+\Gamma_{k-1}U_{k-1} \qquad (6\text{-}4\text{-}8)$$

定义 $k>j$ 时，

$$\Phi_{k/j}=\prod_{i=j}^{k-1}\Phi_{i+1/i}\Phi_{i/i-1}\cdots\Phi_{j+1/j}$$

$$(6\text{-}4\text{-}9)$$

和　　$\Phi_{k/k}=I$

则有

$$X_k=\Phi_{k/0}X_0+\sum_{i=0}^{k-1}\Phi_{k/i+1}B_iU_i+\sum_{i=0}^{k-1}\Phi_{k/i+1}\Gamma_iW_i \qquad (6\text{-}4\text{-}10)$$

(6-4-10) 式表明状态X_k是由初始状态X_0与外加控制U_i、输入噪声W_i等线性组合而成。

§6-5　信号模型的可观测性与可控制性

1. 确定性定常模型的可观测性与可控制性

设有下列信号模型

$$X_{k+1}=\Phi X_k+BU_k \qquad (6\text{-}5\text{-}1)$$

$$Z_{k+1}=HX_{k+1} \qquad (6\text{-}5\text{-}2)$$

我们要问：(1) 在什么条件下，可以通过有限时间长度的测量值，决定出某一时刻的状态矢量，这是可观测性问题。(2) 在什么条件下，可以在有限时间长度内利用外加控制 U使

得信号从初始状态X_0转移到某一终端状态上去，这是可控制性问题。

可观测性

为什么要研究信号模型的可观测性？在回答该问题之前，先来考察一个例子。

例6-5-1 设有线性定常模型为

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{k+1} = \begin{pmatrix} 2 & 0 \\ -1 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_k + \begin{pmatrix} 1 \\ 1 \end{pmatrix} u_k$$

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}_{k+1} = [1, 0] \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{k+1}$$

其中u_k为已知控制量，问状态x_{1k}、x_{2k}是否可观测？

解：因为在i时刻的测量值为

$$z(i) = [1, 0] \begin{pmatrix} x_1(i) \\ x_2(i) \end{pmatrix} = x_1(i)$$

在$i+1$时刻的测量值为

$$z(i+1) = [1, 0] \begin{pmatrix} x_1(i+1) \\ x_2(i+1) \end{pmatrix}$$

$$= [1, 0] \begin{pmatrix} 2 & 0 \\ -1 & -3 \end{pmatrix} \begin{pmatrix} x_1(i) \\ x_2(i) \end{pmatrix} + [1, 0] \begin{pmatrix} 1 \\ 1 \end{pmatrix} u(i)$$

$$= 2x_1(i) + u(i)$$

在$i+2$时刻的测量值为

$$z(i+2) = [1, 0] \begin{pmatrix} x_1(i+2) \\ x_2(i+2) \end{pmatrix}$$

$$= [1, 0] \begin{pmatrix} 2 & 0 \\ -2 & -3 \end{pmatrix}^2 \begin{pmatrix} x_1(i) \\ x_2(i) \end{pmatrix} + [1, 0] \begin{pmatrix} 2 & 0 \\ -1 & -3 \end{pmatrix} u(i) + [1, 0] \begin{pmatrix} 1 \\ 1 \end{pmatrix} u(i+1)$$

$$= 4x_1(i) + 2u(i) + u(i+1)$$

从以上结果可以看到，只要有一个测量值$z(i)$，就能决定出$x_1(i)$，但不论有多少个测量值，也不能决定出$x_2(i)$，所以这个信号模型是不可观测的。从这个例子可以理解到，对于不可观测模型，是不可能从测量值（不管有多少个）确定出状态矢量的，而我们的目的正是希望从测量值中获取某时刻的状态矢量。故这样的信号模型是没有价值的，必需重新构造信号模型，这就回答了为什么要研究模型的可观测性这个问题。

接着我们要问：在什么条件下，才能从测量值决定出状态矢量，也就是在什么条件下，信号模型才是可观测的？在回答该问题之前，也先看一个例子。

例6-5-2 试判断下列模型是否可观测？

$$\begin{pmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & -2 & 1 \\ 3 & 0 & 2 \end{pmatrix} \begin{pmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{pmatrix} + \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} u(k)$$

$$z(k) = [0,\ 1,\ 0] \begin{pmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{pmatrix}$$

解：把模型写成矢量形式，得

$$X_{k+1} = \Phi X_k + BU_k$$

$$Z_k = HX_k$$

在 i 时刻，$Z_i = HX_i$

在 $i+1$ 时刻，
$$Z_{i+1} = HX_{i+1} = H\Phi X_i + HBU_i$$

在 $i+2$ 时刻，
$$Z_{i+2} = HX_{i+2} = H(\Phi X_{i+1} + BU_{i+1}) = H\Phi^2 X_i + H\Phi BU_i + HBU_{i+1}$$

把Φ、B、H代入以上各式，得

$$Z_i = [0,\ 1,\ 0]X_i$$

$$Z_{i+1} = [0,\ -2,\ 1]X_i - U_i$$

$$Z_{i+2} = [3,\ 4,\ 0]X_i + 3U_i - U_{i+1}$$

把 $X_i = \begin{pmatrix} x_{1k} \\ x_{2k} \\ x_{3k} \end{pmatrix}$ 代入以上三式，再合并成矢量形式，结果便得到

$$\begin{pmatrix} z_i \\ z_{i+1} + u_i \\ z_{i+2} - 3u_i + u_{i+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -2 & 1 \\ 3 & 4 & 0 \end{pmatrix} \begin{pmatrix} x_{1i} \\ x_{2i} \\ x_{3i} \end{pmatrix}$$

上式左边是已知矢量，可见，能够解出 i 时刻状态矢量的充分必要条件是矩阵

$$M = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -2 & 1 \\ 3 & 4 & 0 \end{pmatrix}$$

为可逆，或M的秩等于 3（状态矢量维数）。容易检验本例中M的秩等于 3。故本例中模型是可观测的。通过三步测量就可决定出 i 时刻的状态矢量，并且推知，可观测性与外加控制u_i无关。

由于本模型是定常的，上述结论对任意 i 均成立，我们称此模型是完全可观测的。

现在把从例 6-5-2 得到的启示推广到一般情况。因为 BU_k 对模型的可观测性没有影

响，故令$BU_k=0$。对于下列确定性定常线性模型

$$X_{k+1}=\Phi X_k$$

$$Z_{k+1}=H X_{k+1}$$

成为可观测性模型的必要和充分条件是矩阵

$$M=\begin{pmatrix} H \\ H\Phi \\ \vdots \\ H\Phi^{n-1} \end{pmatrix} \tag{6-5-3}$$

的秩等于n（n是状态矢量维数），或矩阵M为可逆。

现在来证明此结论。由于所讨论的模型是定常的，不失一般性，可假定测量是从t_0时刻开始的，要决定的是t_0时刻的状态矢量X_0。记$X_0=[x_{10},\ x_{20},\ \cdots,\ x_{n0}]^T$。因为$U_k$是外加控制量，不失一般性，可令$U_k=0$，那么，$n$个测量值为

$$k=0,\qquad Z_0=HX_0$$

$$k=1,\qquad Z_1=HX_1=H\Phi X_0$$

$$k=2,\qquad Z_2=HX_2=H\Phi^2 X_0$$

……………………

$$k=n-1,\qquad Z_{n-1}=H\Phi^{n-1}X_0$$

把上述方程合并成矢量形式，得

$$\begin{pmatrix} Z_0 \\ Z_1 \\ \vdots \\ Z_{n-1} \end{pmatrix}=\begin{pmatrix} H \\ H\Phi \\ \vdots \\ H\Phi^{n-1} \end{pmatrix}X_0,\qquad X_0=\begin{pmatrix} x_{10} \\ x_{20} \\ \vdots \\ x_{n0} \end{pmatrix} \tag{6-5-4}$$

记

$$Z=\begin{pmatrix} Z_0 \\ Z_1 \\ \vdots \\ Z_{n-1} \end{pmatrix},\qquad M=\begin{pmatrix} H \\ H\Phi \\ \vdots \\ H\Phi^{n-1} \end{pmatrix} \tag{6-5-5}$$

M称为可观测性矩阵。（6-5-4）式中X_0存在唯一解的必要和充分条件是矩阵X可逆，等价的说法是矩阵M满秩，或矩阵的秩等于n。

如果模型仅在i步为可观测，则称此模型在i步可观测，如果对于任意i，模型都是可观测的，则称此模型为完全可观测模型。

对于定常线性模型而言，可观测性与完全可观测性是等同的。

可控制性

所谓可控制性，是状态方程的一种内在属性，就是在输入作用下，状态矢量能不能发生转移。先看一个例子。

例6-5-3　设有状态方程如下

$$\begin{pmatrix} x_1(k+1) \\ x_2(k+1) \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x_1(k) \\ x_2(k) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(k)$$

已知在 $t = 0$ 时刻的状态为 $X(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$，问此模型是否可控制？

解：当 $k = 0$ 时，
$$\begin{pmatrix} x_1(1) \\ x_2(1) \end{pmatrix} = \begin{pmatrix} -1 \\ 2 + u(0) \end{pmatrix}$$

$k = 1$ 时，
$$\begin{pmatrix} x_1(2) \\ x_2(2) \end{pmatrix} = \begin{pmatrix} (-1)^2 \\ 2^2 + 2u(0) + u(1) \end{pmatrix}$$

……………………

$k = k-1$ 时，
$$\begin{pmatrix} u_1(i) \\ u_2(i) \end{pmatrix} = \begin{pmatrix} (-1)^i \\ 2^i + 2^{i-1}u(0) + \cdots + u(k-1) \end{pmatrix}$$

由此可见，第一个状态变量 x_1 永远不受 u 的控制。因此，不能通过输入控制作用使得信号的状态矢量发生转移。一个不能控的状态变量是多余的，应当把它排除掉，并重新构造模型。

上面的例子表明，不是任何信号模型都是可以随意加以控制的，因而有必要研究在什么条件下，模型才是可控制的，在回答上述问题之前，先来考察两个例子。

例6-5-4　给定状态方程为

$$X(k+1) = \Phi X(k) + BU(k)$$

其中
$$\Phi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -2 \\ -1 & 1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

已知初始状态
$$X(0) = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$$

问能否选定控制信号 $u(0)$、$u(1)$、$u(2)$ 使初始状态 $X(0)$ 转移到0状态上去？

解：利用递推法，可得

$$X(1) = \Phi X(0) + BU(0)$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -2 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} u(0)$$

$$= \begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} u(0)$$

$$\boldsymbol{X}(2)=\boldsymbol{\Phi}\boldsymbol{X}(1)+\boldsymbol{B}u(1)$$

$$=\boldsymbol{\Phi}^2\boldsymbol{X}(0)+\boldsymbol{\Phi}\boldsymbol{B}u(0)+\boldsymbol{B}u(1)$$

$$=\begin{pmatrix}2\\6\\0\end{pmatrix}+\begin{pmatrix}1\\-2\\-1\end{pmatrix}u(0)+\begin{pmatrix}1\\0\\1\end{pmatrix}u(1)$$

$$\boldsymbol{X}(3)=\boldsymbol{\Phi}\boldsymbol{X}(2)+\boldsymbol{B}u(2)$$

$$=\boldsymbol{\Phi}^3\boldsymbol{X}(0)+\boldsymbol{\Phi}^2\boldsymbol{B}u(0)+\boldsymbol{\Phi}\boldsymbol{B}u(1)+\boldsymbol{B}u(2)$$

$$=\begin{pmatrix}2\\12\\4\end{pmatrix}+\begin{pmatrix}1\\-2\\-3\end{pmatrix}u(0)+\begin{pmatrix}1\\-2\\-1\end{pmatrix}u(1)+\begin{pmatrix}1\\0\\1\end{pmatrix}u(2)$$

现在令 $\boldsymbol{X}(3)=0$，看能否求出$u(0)$、$u(1)$、$u(2)$，使得从 $\boldsymbol{X}(0)=\begin{pmatrix}2\\1\\0\end{pmatrix}$转移成 $\boldsymbol{X}(3)=\begin{pmatrix}0\\0\\0\end{pmatrix}$。这取决于方程

$$\begin{pmatrix}1\\-2\\-3\end{pmatrix}u(0)+\begin{pmatrix}1\\-2\\-1\end{pmatrix}u(1)+\begin{pmatrix}1\\0\\1\end{pmatrix}u(2)=-\begin{pmatrix}2\\12\\4\end{pmatrix}$$

是否有解。把上述方程改写成

$$\begin{pmatrix}1&1&1\\-2&2&0\\-3&-1&1\end{pmatrix}\begin{pmatrix}u(0)\\u(1)\\u(2)\end{pmatrix}=\begin{pmatrix}-2\\-12\\-4\end{pmatrix}$$

因为系数矩阵 $\begin{pmatrix}1&1&1\\-2&-2&0\\-3&-1&1\end{pmatrix}$的秩等于3，所以有解，其解为

$$\begin{pmatrix}u(0)\\u(1)\\u(2)\end{pmatrix}=\begin{pmatrix}1&1&1\\-2&-2&0\\-3&-1&1\end{pmatrix}^{-1}\begin{pmatrix}-2\\-12\\-4\end{pmatrix}$$

$$
=\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & -1 & \frac{1}{2} \\ 1 & \frac{1}{2} & 0 \end{pmatrix}\begin{pmatrix} -2 \\ -12 \\ -4 \end{pmatrix}=\begin{pmatrix} -5 \\ 11 \\ -8 \end{pmatrix}
$$

所以，如果选取$u(0)=-5$，$u(1)=11$，$u(2)=-8$，就能转移到$X(3)=0$状态上去。

例6-5-5 把上例中的B改为$B=\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$，问能否在第三步上把状态转移到0状态上去？

解： 参照上例，容易算出

$$
X(1)=\begin{pmatrix} 2 \\ 2 \\ -1 \end{pmatrix}+\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}u(0)
$$

$$
X(2)=\begin{pmatrix} 2 \\ 6 \\ 0 \end{pmatrix}+\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}u(0)+\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}u(1)
$$

$$
X(3)=\begin{pmatrix} 2 \\ 12 \\ 4 \end{pmatrix}+\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}u(0)+\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}u(1)+\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}u(2)
$$

令$X(3)=0$，则得到

$$
\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}u(0)-\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}u(1)+\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}u(2)=\begin{pmatrix} -2 \\ -12 \\ -4 \end{pmatrix}
$$

或

$$
\begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} u(0) \\ u(1) \\ u(2) \end{pmatrix}=\begin{pmatrix} -2 \\ -12 \\ -4 \end{pmatrix}
$$

其系数矩阵的秩等于1，是一奇异矩阵，方程组无解。故本例中找不到$u(0)$、$u(1)$、$u(2)$，能使状态$X(0)$转移到0状态上去。还可以推知，即使增加控制步数，模型仍然是不能控制的，因为其系数矩阵不能满秩。

现在把从前面例子得出的启示推广到一般情况。对于可控制性，存在下述结论：

设有确定性线性定常模型的状态方程为

$$
X_{k+1}=\Phi X_k+BU_k
$$

成为能控制性的必要和充分条件是矩阵

$$
C=[B,\ \Phi B,\ \cdots,\ \Phi^{n-1}B] \tag{6-5-6}
$$

的秩等于n，或者说矩阵C满秩，C称为可控制性矩阵。

现在来证明这一结论。因为状态方程

$$X_{k+1} = \Phi X_k + BU_k$$

的解是
$$X_k = \Phi^k X_0 + \sum_{i=0}^{k-1} \Phi^{k-i-1} BU_i \tag{6-5-7}$$

不失一般性，令$X_k = 0$，$k = n$ 则有

$$\Phi^n X_0 + \sum_{i=0}^{n-1} \Phi^{n-i-1} BU_i = 0 \tag{6-5-8}$$

故得
$$X_0 = -\sum_{i=0}^{n-1} \Phi^{-i-1} BU_i \tag{6-5-9}$$

因为Φ是$n \times n$维的，B是$n \times 1$维的，故$\Phi^{-1}B$，$\Phi^{-2}B$，…，$\Phi^{-n}B$都是n维列矢量。展开(6-5-9)式，得

$$X_0 = -[\Phi^{-1}BU_0 + \Phi^{-2}BU_1 + \cdots + \Phi^{-n}BU_{n-1}] \tag{6-5-10}$$

或
$$\Phi^n X_0 = -[\Phi^{n-1}BU_0 + \Phi^{n-2}BU_1 + \cdots + BU_{n-1}] \tag{6-5-11}$$

$$\Phi^n X_0 = -[\Phi^{n-1}B,\ \Phi^{n-2}B, \cdots, B]\begin{pmatrix} U_0 \\ U_1 \\ \vdots \\ U_{n-1} \end{pmatrix} \tag{6-5-12}$$

记
$$C = [\Phi^{n-1}B,\ \Phi^{n-2}B,\ \cdots,\ B] \tag{6-5-13}$$

C称为可控制性矩阵。(6-5-12)式有解的条件是可控制性矩阵可逆，其等价的说法是矩阵C的秩等于n，或矩阵C满秩。

若模型的每一步状态都是可控制的，则称此模型为完全可控制的，若只在第i步上有解，则称此模型为在i步上可控制。对于线性定常模型，可控制性与完全可控制性是等同的，

例6-5-6 已知状态方程

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}_{k+1} = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}_k + \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix}_k$$

问此模型是否可控制？设初始状态为 $X(0) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$，要使得 $X(3) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$，问应该加上的控制$U(0)$、$U(1)$、$U(2)$是何值?

解：检查可控制性矩阵

$$C = [B,\ \Phi B,\ \Phi^2 B]$$

$$= \begin{pmatrix} 1 & 0 & 1 & 2 & 0 & 4 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 4 & 2 \end{pmatrix}$$

其秩等于3，因此本模型是可控制的。

再求控制量$U(0)$、$U(1)$、$U(2)$。我们注意到

$$U(0)=\begin{pmatrix}u_1(0)\\ u_2(0)\end{pmatrix},\quad U(1)=\begin{pmatrix}u_1(1)\\ u_2(1)\end{pmatrix},\quad U(2)=\begin{pmatrix}u_1(2)\\ u_2(2)\end{pmatrix}$$

共有六个未知量，但是$\Phi^n X_0$是三维矢量。由(6-5-12)式，得

$$\Phi^n X_0=-[B,\ \Phi B,\ \Phi^2 B]\begin{pmatrix}u(2)\\ u(1)\\ u(0)\end{pmatrix}$$

把已知的Φ、B、$X(0)$之值代入上式，得

$$-\begin{pmatrix}-12\\ 1\\ 42\end{pmatrix}=\begin{pmatrix}1&0&1&2&0&4\\ 0&1&0&1&0&1\\ 0&0&1&0&4&2\end{pmatrix}\begin{pmatrix}u_1(2)\\ u_2(2)\\ u_1(1)\\ u_2(1)\\ u_1(0)\\ u_2(0)\end{pmatrix}$$

因此这个方程有无穷多组解。换言之，可以有许多种控制量的集合，都能使状态矢量达到预定的转移。

那末，究竟应取那一个控制量集合呢？我们可以加上控制输入能量最小的约束，取其最小范数解(参见附录B)。

对于方程$CU=-\Phi^n X_0$，其最小范数解为

$$U^*=-C^T[CC^T]^{-1}\Phi^n X_0$$

已知 $\Phi^n X_0=\begin{pmatrix}-12\\ 1\\ 42\end{pmatrix}$，$C=\begin{pmatrix}1&0&1&2&0&4\\ 0&1&0&1&0&1\\ 0&0&1&0&4&2\end{pmatrix}$，$CC^T=\begin{pmatrix}22&6&9\\ 6&3&2\\ 9&2&21\end{pmatrix}$。

于是求得最小范数解为

$$U^*=\begin{pmatrix}u_1(2)\\ u_2(2)\\ u_1(1)\\ u_2(1)\\ u_1(0)\\ u_2(0)\end{pmatrix}=\frac{1}{515}\begin{pmatrix}1446\\ -2097\\ -4\\ 795\\ -5800\\ 787\end{pmatrix}$$

2．非定常确定性模型的可观测性与可控制性

假设非定常确定性模型为

$$X_{k+1}=\Phi_{k+1/k}X_k+B_kU_k \tag{6-5-14}$$

$$Z_{k+1}=H_{k+1}X_{k+1} \tag{6-5-15}$$

与定常确定性模型不同之处在于$\Phi_{k+1/k}$、B_k、H_k随时间k而变，状态方程(6-5-14)式的解为

$$X_k = \Phi_{k/0} X_0 + \sum_{i=0}^{k-1} \Phi_{k/i+1} B_i U_i \tag{6-5-16}$$

其中 $\Phi_{k/j} = \Phi_{k/k-1}\Phi_{k-1/k-2}\cdots\Phi_{j+1/j}, \quad j<k$ (6-5-17)

采用与定常确定性模型中类似的方法，可以推导出在t_0时刻的可观测性矩阵为

$$M_0 = \begin{pmatrix} H_0 \\ H_1\Phi_{1/0} \\ \vdots \\ H_{n-1}\Phi_{n-1/0} \end{pmatrix} \tag{6-5-18}$$

若M_0。的秩等于n，说明该模型在t_0时刻的状态X_0是可观测的，但并不说明其他时刻的状态是可观测，因而对任一时刻k，都需要检查其可观测性。

同理，可推导出t_0时刻的可控制性矩阵为

$$C_0 = [B_{n-1},\ \Phi_{n/n-1}B_{n-2},\ \cdots,\ \Phi_{n/2}B_1,\ \Phi_{n/1}B_0] \tag{6-5-19}$$

若C_0。的秩等于n，说明初始状态X_0是能够转移到某设定状态上的，但并不说明其他时刻的状态是可控制，因而对于任一时刻k，都需要检查模型的可控制性。

上面研究了确定模型的可观测性与可控制性条件，这对于研究控制理论无疑是十分重要的，但在滤波理论中，研究随机模型的可观测性与可控制性条件具有更为重要的意义。然而，二者之间有密切联系，从前者容易引伸出后者。

3. 随机定常模型的可观测性

设有定常随机模型为

$$X_k = \Phi X_{k-1} + \Gamma W_{k-1} \tag{6-5-20}$$

$$Z_k = H X_k + V_k \tag{6-5-21}$$

设X_k是n维矢量。

由于模型是定常的，不失一般性，可设测量从零步开始，要决定的是零步状态矢量X_0，其中ΓW_{n-1}是输入噪声。不失一般性，可令$\Gamma=0$。今有n个测量值。

$$\begin{aligned} Z_0 &= H X_0 + V_0 \\ Z_1 &= H X_1 + V_1 = H\Phi X_0 + V_1 \\ Z_2 &= H X_2 + V_2 = H\Phi^2 X_0 + V_2 \\ &\cdots\cdots \\ Z_{n-1} &= H\Phi^{n-1} X_0 + V_{n-1} \end{aligned} \tag{6-5-22}$$

合并成矢量形式，得

$$\begin{pmatrix} Z_0 \\ Z_1 \\ \vdots \\ Z_{n-1} \end{pmatrix} = \begin{pmatrix} H \\ H\Phi \\ \vdots \\ H\Phi^{n-1} \end{pmatrix} X_0 + \begin{pmatrix} V_0 \\ V_1 \\ \vdots \\ V_{n-1} \end{pmatrix} \tag{6-5-23}$$

记

$$Z = \begin{pmatrix} Z_0 \\ Z_1 \\ \vdots \\ Z_{n-1} \end{pmatrix}, \quad M = \begin{pmatrix} H \\ H\Phi \\ \vdots \\ H\Phi^{n-1} \end{pmatrix}, \quad V = \begin{pmatrix} V_0 \\ V_1 \\ \vdots \\ V_{n-1} \end{pmatrix} \tag{6-5-24}$$

则有 $Z = M X_0 + V$

由此得 X_0 的最优加权最小二乘估计为

$$\hat{X}_0 = [M^T R^{-1} M]^{-1} M^T R^{-1} Z \tag{6-5-26}$$

方差阵 $P_0 = [M^T R^{-1} M]^{-1}$ (6-5-27)

其中 $R = E[VV^T]$ (6-5-28)

若 P_0 是正定和有界的，则要求矩阵 M 为可逆，或者说要求矩阵 M 的秩等于 n。

所以，状态 X_0 成为可观测的条件是要求可观测性矩阵

$$M = \begin{pmatrix} H \\ H\Phi \\ \vdots \\ H\Phi^{n-1} \end{pmatrix} \tag{6-5-29}$$

为可逆，或者说要求矩阵 M 的秩等于 n。由于模型是定常的，完全可观测性与可观测性等价。

4. 随机定常模型的可控制性

由(6-4-6)式知状态方程的解为

$$X_n = \Phi^n X_0 + \sum_{i=0}^{n-1} \Phi^{n-i-1} \Gamma W_i$$

设从状态 X_0 出发，经过 n 步转移到达状态 X_n 上去，则有

$$X_n = \Phi^n X_0 + \sum_{i=0}^{n-1} \Phi^{n-i-1} \Gamma W_i \tag{6-5-30}$$

上式可以写成

$$X_n - \Phi^n X_0 = \sum_{i=0}^{n-1} \Phi^{n-i-1} \Gamma W_i$$

$$= [\Phi^{n-1}\Gamma, \ \Phi^{n-2}\Gamma, \cdots, \Gamma] \begin{pmatrix} W_0 \\ W_1 \\ \vdots \\ W_{n-1} \end{pmatrix} \tag{6-5-31}$$

因为模型是定常的，故状态噪声方差为

$E[W_i W_i^T] = Q$，Q 是某常数。这里 X_0、X_n 是随机变量，随机变量 X_0 转移到随机变量 X_n 上去，这种说法是没有意义的，我们只能在均方意义下去研究。

记 $C = E[\Phi^{n-1}\Gamma, \ \Phi^{n-2}\Gamma, \ \cdots, \Phi\Gamma, \ \Gamma]$ (6-5-32)

均方转移误差协方差矩阵为

$$P = E[(X_n - \Phi^n X_0)(X_n - \Phi^n X_0)^T]$$
$$= CQC^T$$

所以，矩阵 P 为正定的条件是要求可控制性矩阵 C 为正定(满秩)。

由于模型是定常的，故完全可控制性条件与可控制性条件等同。

5. 随机非定常模型的可观测性

设有随机非定常模型为

$$\left.\begin{aligned} X_k &= \Phi_{k/k-1} X_{k-1} + \Gamma_{k-1} W_{k-1} \\ Z_k &= H_k X_k + V_k \end{aligned}\right\} \tag{6-5-33}$$

设 X_k是 n 维矢量。由于 $\Gamma_k W_k$ 项对模型的可观测性没有影响，假定 $\Gamma_k W_k=0$，设从 t_0 时刻起，进行 n 次测量，得

$$\left.\begin{aligned} Z_0 &= H_0 X_0 + V_0 \\ Z_1 &= H_1 \Phi_{1/0} X_0 + V_1 \\ &\cdots\cdots\cdots\cdots \\ Z_{n-1} &= H_{n-1} \Phi_{n-1/0} X_0 + V_{n-1} \end{aligned}\right\} \tag{6-5-34}$$

记

$$Z=\begin{pmatrix} Z_0 \\ Z_1 \\ \vdots \\ Z_{n-1} \end{pmatrix},\qquad M=\begin{pmatrix} H_0 \\ H_1\Phi_{1/0} \\ \vdots \\ H_{n-1}\Phi_{n-1/0} \end{pmatrix},\qquad V=\begin{pmatrix} V_0 \\ V_1 \\ \vdots \\ V_{n-1} \end{pmatrix}$$

且有

$$R=E[VV^T]=\begin{pmatrix} R_0 & & 0 \\ & R_1 & \\ & & \ddots \\ 0 & & R_{n-1} \end{pmatrix}$$

把(6-5-34)式写成如下形式

$$Z = M X_0 + V \tag{6-5-35}$$

由(6-5-35)式可求得 X_0 的最优加权最小二乘估计

$$\hat{X}_0 = (M^T R^{-1} M)^{-1} M^T R^{-1} Z \tag{6-5-36}$$

$$P_0 = (M^T R^{-1} M)^{-1} \tag{6-5-37}$$

把P_0^{-1}写成如下形式

$$B = P_0^{-1} = \begin{pmatrix} H_0 \\ H_1\Phi_{1/0} \\ \vdots \\ H_{n-1}\Phi_{n-1/0} \end{pmatrix}^T \begin{pmatrix} R_0^{-1} & & 0 \\ & R_1^{-1} & \\ & & \ddots \\ 0 & & R_{n-1}^{-1} \end{pmatrix} \begin{pmatrix} H_0 \\ H_1\Phi_{1/0} \\ \vdots \\ H_{n-1}\Phi_{n-1/0} \end{pmatrix}$$

$$= \sum_{i=0}^{n-1} \Phi^T_{i/0} H^T_i R^{-1}_i H_i \Phi_{i/0} \tag{6-5-38}$$

若矩阵B可逆，则表明状态 X_0是可观测的，矩阵B称为此模型在 t_0时刻上是可观测性矩阵。

对于任意t_k时刻，可观测性矩阵可表达为

$$B_{k/k-n+1} = \sum_{i=k-n+1}^{k} \Phi^T_{i/k} H_i^T R_i^{-1} H_i \Phi_{i/k} \tag{6-5-39}$$

其中 $\Phi_{i/k}=\Phi_{k/i}^{-1}$，$\Phi_{k/i}=\Phi_{k/k-1}\Phi_{k-1/k-2}\cdots\Phi_{i+1/i}$

若对于任意时刻，模型均是可观测的，则此模型称为完全可观测的。

6. 随机非定常模型的可控制性

对于（6-5-33）式中的状态方程，从初始状态 X_{k-n+1} 出发，施加 n 步状态噪声 W_{k-n}, W_{k-n+1}, …, W_{k-1}, 则有

$$X_k = \Phi_{k/k-n+1} X_{k-n+1} + \sum_{i=k-n+1}^{k} \Phi_{k/i} \Gamma_i W_{i-1} \tag{6-5-40}$$

其中 $\Phi_{k/i} = \Phi_{k/k-1} \Phi_{k-1/k-2} \cdots \Phi_{i+1/i}$

由于模型是非定常的，要使（6-5-40）式成立，即状态 X_{k-n+1} 能够依正概率转移到状态 X_k 上去，则要求矩阵

$$C = E\left[\left(\sum_{i=k-n+1}^{k} \Phi_{k/i} \Gamma_{i-1} W_{i-1}\right)\left(\sum_{i=k-n+1}^{k} \Phi_{k/i} \Gamma_{i-1} W_{i-1}\right)^T\right] \tag{6-5-51}$$

为正定和有界的。

已知 W_i 为零均值白噪声，且 $E[W_i W_i^T] = Q_i$，则（6-5-41）式可化成

$$C = \sum_{i=k-n+1}^{k} \Phi_{k/i} \Gamma_{i-1} Q_{i-1} \Gamma_{i-1}^T \Phi_{k/i}^T \tag{6-5-42}$$

矩阵 C 称为非定常随机模型的可控制性矩阵。若矩阵 C 是正定和有界的，表明从状态 X_{k-n+1} 出发可以依正概率转移到状态 X_k 上去。若对于任意时刻，模型都是可控制的，称此模型为完全可控制的。

例6-5-1　试检验 $\alpha-\beta-\gamma$ 滤波器中状态方程的可观测性与可控制性。

已知此信号模型为

$$\begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \end{pmatrix}_k = \begin{pmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \end{pmatrix}_{k-1} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} w_{k-1}$$

$$z_k = [1\ 0\ 0] \begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \end{pmatrix} + v_k$$

$$E[w_k^2] = Q, \qquad E[v_k^2] = R$$

解：显见此模型为定常随机模型，记

$$\Phi = \begin{pmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & 1 \end{pmatrix} \qquad \Gamma = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad H = [1,\ 0,\ 0]$$

代入可观测性矩阵，得

$$\Phi^2 = \begin{pmatrix} 1 & 2T & 2T^2 \\ 0 & 1 & 2T \\ 0 & 0 & 1 \end{pmatrix} \qquad M = \begin{pmatrix} H \\ H\Phi \\ H\Phi^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & T & \frac{T^2}{2} \\ 1 & 2T & 2T^2 \end{pmatrix}$$

M 的秩等于 3，M 为满秩，所以本模型是可观测的，故由测量值 Z_k、Z_{k+1}、Z_{k+2} 可确定出状态估计 $\hat{X}_k$。

再代入可控制性矩阵，得

$$C=[\Gamma,\ \Phi\Gamma,\ \Phi^2\Gamma]$$

$$=\begin{pmatrix} 0 & T^2/2 & 2T^2 \\ 0 & T & 2T \\ 1 & 1 & 1 \end{pmatrix}$$

C的秩等于3，C为满秩。所以本模型是可控制的。故可以在输入噪声ΓW_k、ΓW_{k+1}、ΓW_{k+2}的作用下，从初始状态X_k以正的概率转移到终端状态X_{k+3}上去。

若在上例中，状态方程保持不变，两测量值改为速度，则测量方程为

$$z_k=[0,\ 1,\ 0]\begin{pmatrix} x_k \\ \dot{x}_k \\ \ddot{x}_k \end{pmatrix}+\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}v_k$$

试检验此模型的可观测性与可控制性。

由可观测性矩阵得

$$M=\begin{pmatrix} H \\ H\Phi \\ H\Phi^2 \end{pmatrix}=\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & T \\ 0 & 1 & 2T \end{pmatrix}$$

显见矩阵M的秩小于3，M是奇异矩阵。

再由可控制性矩阵，得

$$C=[\Gamma,\ \Phi\Gamma,\ \Phi^2\Gamma]$$

$$=\begin{pmatrix} 0 & \dfrac{T^2}{2} & 2T^2 \\ 0 & T & 2T \\ 1 & 1 & 1 \end{pmatrix}$$

显见矩阵C的秩等于3。

所以，此模型是不可观测的，但是可控制的。因而不能从观测值得到状态矢量的估计。

第七章　离散卡尔曼滤波

从维纳滤波器理论知道，对于非平稳随机过程，维纳滤波是难以实现的，到六十年代初，卡尔曼把状态变量法应用于过程估计中。卡尔曼滤波理论的实质是递归滤波的方法，采用线性无偏最小方差准则以得到过程的最优估计，可以应用于多输入、多输出非平稳随机过程，这在估计理论上是一大突破。以后的实践证明，卡尔曼滤波在许多领域有着十分广泛的应用。

本章着重研究离散线性卡尔曼滤波的各种性质。

§7-1　离散卡尔曼滤波

设有一可观测的和可控制的线性过程，其状态方程和测量方程为

$$\left.\begin{aligned} X_{k+1} &= \Phi_{k+1/k} X_k + W_k \\ Z_{k+1} &= H_{k+1} X_{k+1} + V_{k+1} \end{aligned}\right\} \tag{7-1-1}$$

其中W_k和V_k的统计性质为

$$E[W_k]=0,\qquad E[W_k W_j^T]=Q_k\delta_{kj}$$

$$E[V_k]=0,\qquad E[V_k V_j^T]=R_k\delta_{kj}$$

$$E[W_k V_j^T]=0$$

其中　$\delta_{kj}=\begin{cases}1 & k=j\\ 0 & k\neq j\end{cases}$

给定初始条件为

$$\hat{X}_0=E[X_0]=\overline{X}_0$$

$$P_0=E[(X_0-\overline{X}_0)(X_0-\overline{X}_0)^T]=\mathrm{var}\,X_0$$

$$\mathrm{cov}[X_0,W_k]=0\qquad \mathrm{cov}[X_0,V_k]=0$$

根据以上条件，寻求X_{k+1}的最优估计。

估计分两步进行，第一步，假定已得到k时刻的估计$\hat{X}_k$，根据状态方程可以作出对$k+1$时刻的预报估计

$$\hat{X}_{k+1/k}=\Phi_{k+1/k}\hat{X}_k \tag{7-1-2}$$

若$\hat{X}_k$是X_k的无偏估计，则有

$$E[\hat{X}_{k+1/k}]=E[\Phi_{k+1/k}\hat{X}_k]=\Phi_{k+1/k}E[\hat{X}_k]=\Phi_{k+1/k}\overline{X}_k=E[X_{k+1}] \tag{7-1-3}$$

显然，$\hat{X}_{k+1/k}$是X_{k+1}的一个无偏估计。

预报估计误差为

$$\tilde{X}_{k+1/k}=\hat{X}_{k+1/k}-X_{k+1}$$
$$=\hat{X}_{k+1/k}-\Phi_{k+1/k}X_k-W_k \tag{7-1-4}$$

由此得预报估计误差协方差阵为

$$P_{k+1/k}=E[\tilde{X}_{k+1/k}\tilde{X}^T_{k+1/k}]$$
$$=E[(\Phi_{k+1/k}\hat{X}_k-\Phi_{k+1/k}X_k-W_k)(\Phi_{k+1/k}\hat{X}_k-\Phi_{k+1/k}X_k-W_k)^T]$$
$$=E[(\Phi_{k+1/k}\tilde{X}_k-W_k)(\Phi_{k+1/k}\tilde{X}_k-W_k)^T]$$

其中$\tilde{X}_k=\hat{X}_k-X_k$，再利用关系$E[\tilde{X}_kW^T_k]=E[W_k\tilde{X}^T_k]=0$，便得出

$$P_{k+1/k}=\Phi_{k+1/k}P_k\Phi^T_{k+1/k}+Q_k \tag{7-1-5}$$

第二步，根据$k+1$时刻的新测量值Z_{k+1}，对X_{k+1}作第二个估计$\hat{X}'_{k+1}$

$$Z_{k+1}=H_{k+1}\hat{X}'_{k+1} \tag{7-1-6}$$

因为$E[Z_{k+1}]=H_{k+1}E[X_{k+1}]=H_{k+1}E[\hat{X}'_{k+1}]$，故$\hat{X}'_{k+1}$是$X_{k+1}$的又一无偏估计。

把两个无偏估计$\hat{X}_{k+1/k}$和$\hat{X}'_{k+1}$进行线性组合，令

$$\hat{X}_{k+1}=(1-B_{k+1})\hat{X}_{k+1/k}+B_{k+1}\hat{X}'_{k+1} \tag{7-1-7}$$

其中B_{k+1}为加权系数，再令

$$B_{k+1}=K_{k+1}H_{k+1}$$

代入(7-1-7)式，得

$$\hat{X}_{k+1}=(1-K_{k+1}H_{k+1})\hat{X}_{k+1/k}+K_{k+1}Z_{k+1} \tag{7-1-8}$$
$$=\hat{X}_{k+1/k}+K_{k+1}[Z_{k+1}-H_{k+1}\hat{X}_{k+1/k}]$$

其中K_{k+1}为未知权系数矩阵，现在的问题是要根据某种最优准则选取适当的K_{k+1}，以得到最优估计$\hat{X}_{k+1}$。

现在仍然采用均方误差最小准则，要求均方误差矩阵的迹达到最小，即使下式达到最小

$$J_{k+1}=\mathrm{tr}E[(\hat{X}_{k+1}-X_{k+1})(\hat{X}_{k+1}-X_{k+1})^T] \tag{7-1-9}$$

因为$\hat{X}_{k+1}$是两个无偏估计的线性组合，不难证明$\hat{X}_{k+1}$也是X_{k+1}的无偏估计。对于无偏估计而言，均方误差最小等价于方差最小，令估计误差为

$$\tilde{X}_{k+1}=X_{k+1}-\hat{X}_{k+1} \tag{7-1-10}$$

把$\hat{X}_{k+1}$和X_{k+1}的表达式代入，得

$$\tilde{X}_{k+1}=X_{k+1}-K_{k+1}H_{k+1}X_{k+1}+K_{k+1}H_{k+1}Z_{k+1}-(1-K_{k+1}H_{k+1})\hat{X}_{k+1/k}-K_{k+1}Z_{k+1}$$
$$=(1-K_{k+1}H_{k+1})(X_{k+1}-\hat{X}_{k+1/k})+K_{k+1}(H_{k+1}X_{k+1}-Z_{k+1})$$
$$=(1-K_{k+1}H_{k+1})\tilde{X}_{k+1/k}-K_{k+1}V_{k+1} \tag{7-1-11}$$

故估计误差的协方差阵为

$$P_{k+1}=E[\tilde{X}_{k+1}\tilde{X}^T_{k+1}]$$
$$=(1-K_{k+1}H_{k+1})P_{k+1/k}(1-K_{k+1}H_{k+1})^T+K_{k+1}R_{k+1}K^T_{k+1} \tag{7-1-12}$$

现在来求$\mathrm{tr}P_{k+1}$的极小值，令$\frac{\partial}{\partial K_{k+1}}\mathrm{tr}P_{k+1}=0$

$$\mathrm{tr}P_{k+1}=\mathrm{tr}\{P_{k+1/k}-K_{k+1}H_{k+1}P_{k+1/k}-P_{k+1/k}H_{k+1}^{T}K_{k+1}^{T}$$
$$-K_{k+1}H_{k+1}P_{k+1}H_{k+1}^{T}K_{k+1}^{T}+K_{k+1}R_{k+1}K_{k+1}^{T}\} \tag{7-1-13}$$

利用附录中矩阵迹求导数公式，得

$$\frac{\partial}{\partial K_{k+1}}\mathrm{tr}P_{k+1}=-2P_{k+1/k}H_{k+1}^{T}+2K_{k+1}H_{k+1}P_{k+1/k}H_{k+1}^{T}+2K_{k+1}R_{k+1}=0 \tag{7-1-14}$$

解出

$$K_{k+1}=P_{k+1/k}H_{k+1}^{T}[H_{k+1}P_{k+1/k}H_{k+1}^{T}+R_{k+1}]^{-1} \tag{7-1-15}$$

上式中的K_{k+1}称为最佳权矩阵，或称为卡尔曼滤波器的最佳增益矩阵。在最佳增益条件下，$\mathrm{tr}P_{k+1}$达到最小，此时P_{k+1}称为估计误差的最小协方差阵。把(7-1-15)式代入(7-1-12)式中，得

$$P_{k+1}=(1-K_{k+1}H_{k+1})P_{k+1/k} \tag{7-1-16}$$

把以上结果进行整理，便得到卡尔曼滤波器的一组递推公式：

预报估计

$$\hat{X}_{k+1/k}=\Phi_{k+1/k}\hat{X}_{k} \tag{7-1-17}$$

预报误差协方差阵

$$P_{k+1/k}=\Phi_{k+1/k}P_{k}\Phi_{k+1/k}^{T}+Q_{k} \tag{7-1-18}$$

在得到新测量值Z_{k+1}之后，进行估计更新，得滤波估计

$$\hat{X}_{k+1}=\hat{X}_{k+1/k}+K_{k+1}[Z_{k+1}-H_{k+1}\hat{X}_{k+1/k}] \tag{7-1-19}$$

滤波误差协方差阵

$$P_{k+1}=(1-K_{k+1}H_{k+1})P_{k+1/k} \tag{7-1-20}$$

最佳增益

$$K_{k+1}=P_{k+1/k}H_{k+1}^{T}[H_{k+1}P_{k+1/k}H_{k+1}^{T}+R_{k+1}]^{-1} \tag{7-1-21}$$

有了这一组递推公式之后，如果已知k时刻的估计$\hat{X}_k$及估计误差的协方差矩阵P_k就可以根据$k+1$时刻的测量值Z_{k+1}，得到$k+1$时刻的估计$\hat{X}_{k+1}$与相应的P_{k+1}。

递推计算从$k=0$时刻开始，为了启动卡尔曼滤波器，需要给出初始值$\hat{X}_0$和P_0；为了使滤波器的输出每一步都成为线性无偏最小方差估计，要求初值满足下列条件

$$\hat{X}_0=E[X_0]=\overline{X}_0 \tag{7-1-22}$$

$$P_0=E[(X_0-\overline{X}_0)(X_0-\overline{X}_0)^{T}]=X_0 \tag{7-1-23}$$

卡尔曼滤波的等价形式

(7-1-17)～(7-1-21)五式构成了卡尔曼滤波的递推形式。经过适当运算可以得到某些等价形式。

已知最佳增益

$$K_{k+1}=P_{k+1/k}H_{k+1}^{T}(H_{k+1}P_{k+1/k}H_{k+1}^{T}+R_{k+1})^{-1} \tag{7-1-24}$$

按照附录D中矩阵等式，得

$$K_{k+1}=(P_{k+1/k}^{-1}+H_{k+1}^{T}R_{k+1}^{-1}H_{k+1})^{-1}H_{k+1}^{T}R_{k+1}^{-1} \tag{7-1-25}$$

把(7-1-25)式代入P_{k+1}表达式中，得

$$P_{k+1}=(1-K_{k+1}H_{k+1})P_{k+1/k}$$

$$
\begin{aligned}
P_{k+1} &= [I-(P^{-1}_{k+1/k}+H^T_{k+1}R^{-1}_{k+1}H_{k+1})^{-1}H^T_{k+1}R^{-1}_{k+1}H_{k+1}]P_{k+1/k} \\
&= (P^{-1}_{k+1/k}+H^T_{k+1}R^{-1}_{k+1}H_{k+1})^{-1}[P^{-1}_{k+1/k}+H^T_{k+1}R^{-1}_{k+1}H_{k+1}-H^T_{k+1}R^{-1}_{k+1}H_{k+1}]P_{k+1/k} \\
&= (P^{-1}_{k+1/k}+H^T_{k+1}R^{-1}_{k+1}H_{k+1})^{-1} \qquad (7\text{-}1\text{-}26)
\end{aligned}
$$

把(7-1-26)式代回(7-1-25)式中，得

$$K_{k+1}=P_{k+1}H^T_{k+1}R^{-1}_{k+1} \qquad (7\text{-}1\text{-}27)$$

现在，重新整理一下卡尔曼滤波器公式，结果如下：

预报估计

$$\hat{X}_{k+1}=\Phi_{k+1/k}\hat{X}_k \qquad (7\text{-}1\text{-}28)$$

预报误差协方差阵

$$P_{k+1/k}=\Phi_{k+1/k}P_k\Phi^T_{k+1/k}+Q_k \qquad (7\text{-}1\text{-}29)$$

最佳增益

$$
\begin{aligned}
K_{k+1} &= P_{k+1/k}H^T_{k+1}(H_{k+1}P_{k+1/k}H^T_{k+1}+R_{k+1})^{-1} \qquad (7\text{-}1\text{-}30a) \\
&= (P^{-1}_{k+1/k}+H^T_{k+1}R^{-1}_{k+1}H_{k+1})^{-1}H^T_{k+1}R^{-1}_{k+1} \qquad (7\text{-}1\text{-}30b) \\
&= P_{k+1}H^T_{k+1}R^{-1}_{k+1} \qquad (7\text{-}1\text{-}30c)
\end{aligned}
$$

滤波估计

$$\hat{X}_{k+1}=\hat{X}_{k+1/k}+K_{k+1}[Z_{k+1}-H_{k+1}\hat{X}_{k+1/k}] \qquad (7\text{-}1\text{-}31)$$

滤波误差协方差阵

$$
\begin{aligned}
P_{k+1} &= (I-K_{k+1}H_{k+1})P_{k+1/k}(I-K_{k+1}H_{k+1})^T+K_{k+1}R_{k+1}K^T_{k+1} \qquad (7\text{-}1\text{-}32a) \\
&= (I-K_{k+1}H_{k+1})P_{k+1/k} \qquad (7\text{-}1\text{-}32b) \\
&= (P^{-1}_{k+1/k}+H^T_{k+1}R^{-1}_{k+1}H_{k+1})^{-1} \qquad (7\text{-}1\text{-}32c)
\end{aligned}
$$

最佳初值 $\hat{X}_0=\bar{X}_0$， $P_0=\mathrm{var}\,X_0$

因为各种形式都是等价的，可以根据计算上的方便任意选用一种形式。

下面对卡尔曼滤波器的性质作几点说明：

1. 在卡尔曼滤波理论中，把状态矢量和测量矢量看成是由白噪声序列驱动一个线性滤波器所产生。根据描述信号过程的数学模型

$$
\left.
\begin{aligned}
X_{k+1} &= \Phi_{k+1/k}X_k+\Gamma_kW_k \\
Z_{k+1} &= H_{k+1}X_{k+1}+V_{k+1}
\end{aligned}
\right\} \qquad (7\text{-}1\text{-}33)
$$

可以画出信号形成器框图，如图7-1-1所示。

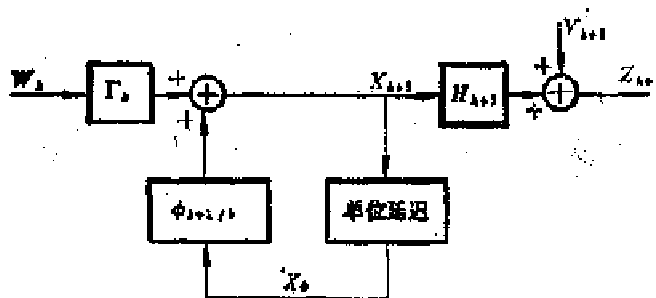


图7-1-1 信号形成器框图

卡尔曼滤波器就是要从测量序列得到状态矢量的估计，根据滤波方程

$$\hat{X}_{k+1}=\Phi_{k+1/k}\hat{X}_k+K_{k+1}[Z_{k+1}-H_{k+1}\Phi_{k+1/k}\hat{X}_k]$$

可以画出卡尔曼滤波器框图，如图7-1-2所示。可见一定的信号模型与一定的卡尔曼滤波器对应。

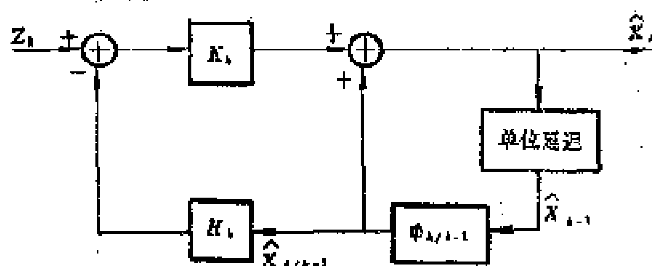


图7-1-2 卡尔曼滤波器框图

2. 应用上述卡尔曼滤波时，要求W_k、V_k均为零均值白噪声，且互不相关，并精确知道协方差阵Q_k和R_k之值。若这些先验知识知道得不准确，那么，算出的权矩阵K_k就不是最佳的，得出的估计$\hat{X}_k$也不是无偏和最小方差的。

3. 要求给出正确的初始值为

$$\hat{X}_0 = E[X_0] = \overline{X}_0$$
$$P_0 = E[(X_0 - \overline{X}_0)(X_0 - \overline{X}_0)^T] = \text{var}\, X_0$$

在给定正确的初值条件下，得出的每一步滤波估计才是无偏的和最小方差的，否则就不是。幸好，卡尔曼滤波器有一个优越的性质：在迭代时间充分长之后，不正确初值的影响会逐渐消失，滤波器将收敛到最佳状态。关于卡尔曼滤波器的收敛性质留在以后说明。

4. 尔卡曼滤波器类似于一反馈调节机构，具有自较正作用。如果状态噪声强度Q_k很小，那么一步预报$\hat{X}_{k+1/k}$很准确，测量值Z_{k+1}的意义就很小，增益K_{k+1}就应当小。反之如果测量噪声强度R_{k+1}很小，意味着测量很准确，预报的意义就很小，增益K_{k+1}就应增大因此卡尔曼滤波器中的增益自动地调节到最佳，故卡尔曼滤波器是时变参数滤波器，这也是卡尔曼滤波器优于一切时不变滤波器的原因所在。

5. 由于最佳增益决定于P_0、Q_k和R_k，而与测量值和估计值无关，因此，各个最佳增益可以事先计算好，有利于提高计算速度。

卡尔曼滤波的计算流程与应用举例

在给定初值$\hat{X}_0$和P_0，噪声强度Q_k和R_k之后，就可递推地计算各时刻的估计值及估计误差的协方差矩阵。卡尔曼滤波的计算流程如图7-1-3所示。

下面举两个例子：

例7-1-1 设动态过程的状态矢量为$X_k = \begin{bmatrix} x_{1k} \\ x_{2k} \end{bmatrix}$，状态方程为$X_k = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \times X_{k-1} + W_{k-1}$，测量方程为$Z_k = [1,0] \times \begin{bmatrix} x_{1k} \\ x_{2k} \end{bmatrix} + V_k$

已知$E[W_k] = 0, E[W_k W_k^T] = Q_k = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, E[V_k] = 0, E[V_k V_k^T] = R_k = 2 + (-1)^k$。

这就是说，当时间k为偶数时的测量噪声要比k为奇数时的测量噪声强些，初值$P_0 = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$，求各时刻的滤波器参数$P_{k/k-1}$、$K_k$、$P_k$。

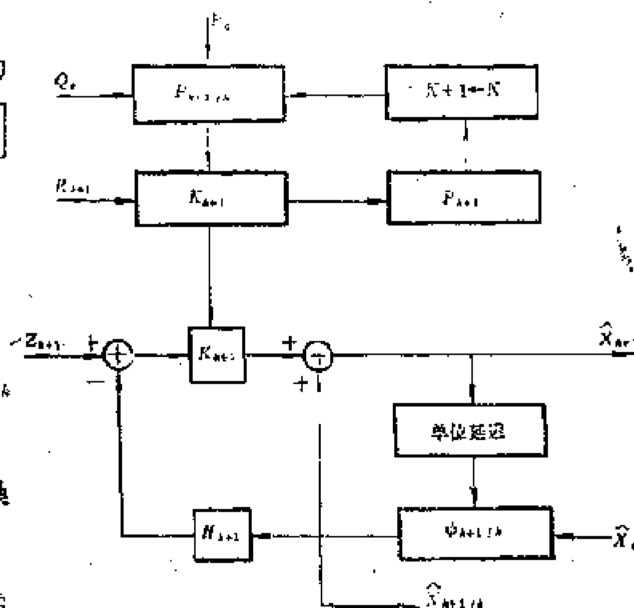


图7-1-3 卡尔曼滤波器计算流程图

解：代人卡尔曼滤波公式中，得

$$P_{1/0}=\Phi_{1/0}P_0\Phi_{1/0}^T+Q_0$$

$$=\begin{bmatrix}1&1\\0&1\end{bmatrix}\begin{bmatrix}10&0\\0&10\end{bmatrix}\begin{bmatrix}1&0\\1&1\end{bmatrix}+\begin{bmatrix}0&0\\0&1\end{bmatrix}=\begin{bmatrix}20&11\\10&11\end{bmatrix}$$

$$K_1=\begin{bmatrix}20&10\\10&11\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}\left\{[1,0]\begin{bmatrix}20&10\\10&11\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}+2+(-1)\right\}^{-1}$$

$$=\frac{1}{21}\begin{bmatrix}20\\10\end{bmatrix}=\begin{bmatrix}0.95\\0.48\end{bmatrix}$$

$$P_1=\left\{\begin{bmatrix}1&0\\0&1\end{bmatrix}-\frac{1}{21}\begin{bmatrix}20\\10\end{bmatrix}[1,\ 0]\right\}\begin{bmatrix}20&10\\10&11\end{bmatrix}$$

$$=\frac{1}{2}\begin{bmatrix}20&10\\10&21\end{bmatrix}=\begin{bmatrix}0.95&0.48\\0.48&0.24\end{bmatrix}$$

依此类推，计算了六步，所得结果如下：

表7-1-1　例 7-1-1 计 算 结 果

K	$P_{k/k-1}$	K_k	P_k
0			$\begin{bmatrix}10&0\\0&10\end{bmatrix}$
1	$\begin{bmatrix}20&10\\10&21\end{bmatrix}$	$\begin{bmatrix}0.95\\0.48\end{bmatrix}$	$\begin{bmatrix}0.95&0.48\\0.48&0.24\end{bmatrix}$
2	$\begin{bmatrix}8.14&0.71\\0.71&7.24\end{bmatrix}$	$\begin{bmatrix}0.73\\0.6\end{bmatrix}$	$\begin{bmatrix}2.19&1.82\\1.82&3.21\end{bmatrix}$
3	$\begin{bmatrix}9.05&5\\5&4.21\end{bmatrix}$	$\begin{bmatrix}0.9\\0.5\end{bmatrix}$	$\begin{bmatrix}0.9&0.5\\0.5&1.7\end{bmatrix}$
4	$\begin{bmatrix}3.6&2.2\\2.2&4.21\end{bmatrix}$	$\begin{bmatrix}0.55\\0.33\end{bmatrix}$	$\begin{bmatrix}1.02&1\\1&2\end{bmatrix}$
5	$\begin{bmatrix}5.6&3\\3&2\end{bmatrix}$	$\begin{bmatrix}0.85\\0.45\end{bmatrix}$	$\begin{bmatrix}0.85&0.45\\0.45&0.63\end{bmatrix}$

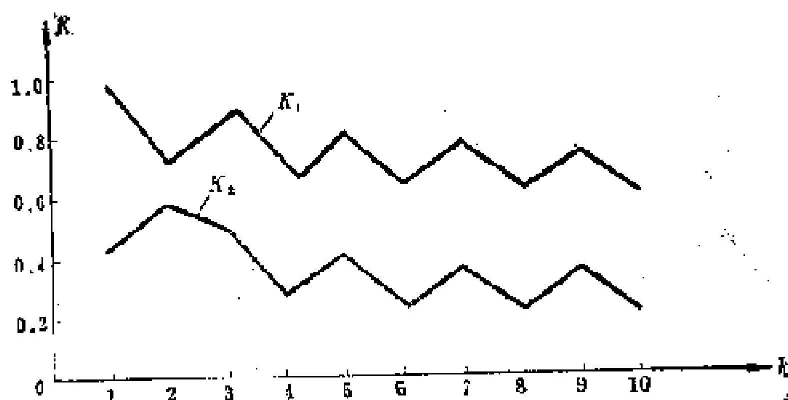


图7-1-4　例7-1-1的计算结果

从图7-1-4可以看到，当k为奇数时的增益要比k为偶数时的增益大些，这是因为k为奇数时，测量噪声较弱，测量值较可靠，故对测量值较倚重。此外，我们还注意到，随着k的增大，各增益值都趋于常数，即使重新取一个P_0，我们发现各增益仍然趋于原来的常

数，不仅如此，协方差阵P_k的各元素也都趋于常数，这种现象称为卡尔曼滤波的收敛特性，这是一个十分重要和有用的性质。

根据算出的K_k值，给出初值$\hat{X}_0$和各次测量值Z_k，就可以算出各时刻的状态估计$\hat{X}_k$

例7-1-2　用卡尔曼滤波器求落体的高度和速度的估计。测量值是落体的高度，给出初始估计$\hat{X}_0=95\text{m}$，$\hat{\dot{X}}_0=1\text{m/sec}$，初始方差为$\text{var}\,X_0=10\text{m}^2$，$\text{var}\,\dot{X}_0=1(\text{m/sec})^2$，测量间隔 $T=1\text{sec}$，已知六次测量值 $Z_1=100$ m，$Z_2=97.9$ m，$Z_3=94.4$ m，$Z_4=92.7$ m，$Z_5=87.3\text{m}$，$Z_6=82.1\text{m}$。

解：因为落体的运动方程为

$$\ddot{x}(t)=-g \qquad t\geqslant 0$$

g为加速度，等于 1 m/sec²，今用$x_1(t)$表示t时刻落体高度，用 $x_2(t)$ 表示t时刻落体速度。

则有
$$\begin{cases}\dot{x}_1(t)=x_2(t)\\ \dot{x}_2(t)=-g\end{cases}$$

写成矢量形式，得

$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{bmatrix}=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix}+\begin{bmatrix}0\\ -g\end{bmatrix}$$

令
$$X(t)=\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix},\quad A=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix},\quad U=\begin{bmatrix}0\\ -g\end{bmatrix}$$

则有
$$\dot{X}(t)=AX(t)+U$$

其解为
$$X(t)=e^{A(t-t_0)}X(t_0)+\int_{t_0}^{t}e^{A(t-t_0)}U\,dt$$

令
$$t_0=kT,\quad t=(k+1)T$$

则
$$e^{(At-t_0)}=e^{AT}=e^{\left[\begin{smallmatrix}0 & T\\ 0 & 0\end{smallmatrix}\right]}\approx 1+AT=\begin{bmatrix}1 & T\\ 0 & 1\end{bmatrix}$$

再令
$$t_0=kT,\quad t=kT+\tau$$

则
$$\int_{t_0}^{t}e^{A(t-t_0)}U\,dt=\int_0^T e^{A\tau}U\,d\tau=\int_0^T\begin{bmatrix}1 & \tau\\ 0 & 1\end{bmatrix}\begin{bmatrix}0\\ -g\end{bmatrix}d\tau=\int_0^T\begin{bmatrix}-g\tau\\ -g\end{bmatrix}d\tau$$

$$=-\begin{bmatrix}\dfrac{gT^2}{2}\\ gT\end{bmatrix}$$

于是得到落体的离散时间状态方程为

$$X_{k+1}=\begin{bmatrix}1 & T\\ 0 & 1\end{bmatrix}X_k+\begin{bmatrix}\dfrac{T^2}{2}\\ T\end{bmatrix}(-g)$$

或
$$X_{k+1}=\Phi_{k+1/k}X_k+B_k u_k$$

其中 $\Phi_{k+1/k}=\begin{bmatrix}1 & T\\0 & 1\end{bmatrix}$，$B_k=\begin{bmatrix}\dfrac{T^2}{2}\\T\end{bmatrix}$，$u_k=-g$

已知测量方程为

$$Z_k=[1\quad 0]X_k+V_k$$

这相当于有测量噪声而没有状态噪声的情况，但有外加控制($-g$)。

根据给定的初值和卡尔曼滤波公式，可以递推地计算出各时刻的落体的高度和速度的估计，为使计算简化，令$g=1\text{m/sec}^2$，得出

$$\hat{X}_{1/0}=\begin{bmatrix}0 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}95\\1\end{bmatrix}-\begin{bmatrix}0.5\\1.0\end{bmatrix}=\begin{bmatrix}95.5\\0\end{bmatrix}$$

$$P_{1/0}=\begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}\begin{bmatrix}10 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\1 & 1\end{bmatrix}=\begin{bmatrix}11 & 1\\1 & 1\end{bmatrix}$$

$$K_1=\begin{bmatrix}11 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}\left\{[1,\ 1]\begin{bmatrix}11 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}+1\right\}^{-1}=\begin{bmatrix}0.92\\0.08\end{bmatrix}$$

$$\hat{X}_1=\begin{bmatrix}95.5\\0\end{bmatrix}+\begin{bmatrix}0.92\\0.08\end{bmatrix}(100-95.5)=\begin{bmatrix}99.64\\0.36\end{bmatrix}$$

$$P_1=\left\{\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}-\begin{bmatrix}0.92\\0.08\end{bmatrix}[1,\ 0]\right\}\begin{bmatrix}11 & 1\\1 & 1\end{bmatrix}=\begin{bmatrix}0.88 & 0.10\\0.10 & 0.92\end{bmatrix}$$

到此，得到了第一次估计$\hat{X}_1$和P_1依此类推，可得出各个时刻的估计$\hat{X}_k$和P_k。

表7-1-2给出全部计算成果，表中还给出了落体的高度和速度的真值，以资比较。由表中数据可看到，当递推至$k=6$时，估计值与真值已相当接近，估计误差的方差也越来越小。

表7-1-2

k	0	1	2	3	4	5	6
X_k(m)	100.0	99.5	98.0	95.5	92.0	87.5	82.0
$\dot{X}_k$(m/sec)	0	−1.0	−2.0	−3.0	−4.0	−5.0	−6.0
V_k(m)		0.5	−0.1	1.1	0.7	−0.2	0.1
Z_k(m)		100.0	97.9	94.4	92.7	87.3	82.1
$\hat{X}_k$(m)	95.0	99.6	98.4	95.3	92.6	87.9	82.4
$\hat{\dot{X}}_k$(m/sec)	1.0	0.36	−1.2	−2.3	−3.25	−4.6	−5.7
$\sigma^2_{\hat{x}}$(m²)	10.0	0.88	0.66	0.66	0.62	0.56	0.50
$\sigma^2_{\hat{\dot{x}}}$(m²/sec²)	1.0	0.92	0.57	0.30	0.16	0.10	0.06

具有已知外加控制和测量偏差时的卡尔曼滤波算法

由于外加控制和测量偏差都是已知值，我们仅需对(7-1-1)式略作修改就得出相应的状态方程和测量方程

$$X_{k+1}=\Phi_{k+1/k}X_k+B_kU_k+\Gamma_kW_k \tag{7-1-34}$$

$$Z_{k+1}=H_{k+1}X_{k+1}+b_{k+1}+V_{k+1} \tag{7-1-35}$$

其中U_k为外加控制，B_k为控制转移矩阵，b_{k+1}为测量偏差。由于没有改变噪声的性质，所以只需对原来的卡尔曼滤波器略作修改就得出相应的卡尔曼滤波器如下：

预报估计

$$\hat{X}_{k+1/k}=\Phi_{k+1/k}\hat{X}_k+B_kU_k \tag{7-1-36}$$

预报误差协方差阵

$$P_{k+1/k}=\Phi_{k+1/k}P_k\Phi^T_{k+1/k}+\Gamma_kQ_k\Gamma^T_k \tag{7-1-37}$$

最佳增益

$$K_{k+1}=P_{k+1/k}H^T_{k+1}[H_{k+1}P_{k+1/k}H^T_{k+1}+R_{k+1}]^{-1} \tag{7-1-38}$$

滤波估计

$$\hat{X}_{k+1}=\hat{X}_{k+1/k}+K_{k+1}[Z_{k+1}-b_{k+1}-H_{k+1}\hat{X}_{k+1/k}] \tag{7-1-39}$$

滤波误差协方差阵

$$P_{k+1}=(I-K_{k+1}H_{k+1})P_{k+1/k} \tag{7-1-40}$$

状态噪声与测量噪声相关时的卡尔曼滤波器

设动态过程的数学模型为

$$X_{k+1}=\Phi_{k+1/k}X_k+\Gamma_kW_k \tag{7-1-41}$$

$$Z_{k+1}=H_{k+1}X_{k+1}+V_{k+1} \tag{7-1-42}$$

式中W_k、V_k仍为白噪声序列，但二者相关。

$$E\left\{\begin{bmatrix}W_k\\V_k\end{bmatrix}[W_j^T\quad V_j^T]\right\}=\begin{bmatrix}Q_k & S_k\\S_k^T & R_k\end{bmatrix}\delta_{kj} \tag{7-1-43}$$

例如，用陀螺仪作为指北向的基准，当载体作急速转弯时，转弯加速度使陀螺产生漂移，使得测量噪声与状态噪声发生相关。

令
$$\overline{W}_k=W_k-S_kR_k^{-1}V_k \tag{7-1-44}$$

则有
$$E\left\{\begin{bmatrix}\overline{W}_k\\V_k\end{bmatrix}[\overline{W}_k^T\quad V_k^T]\right\}=\begin{bmatrix}Q_k-S_kR_k^{-1}S_k^T & 0\\0 & R_k\end{bmatrix} \tag{7-1-45}$$

可见$\overline{W}_k$与V_k不相关。

重写状态方程

$$\begin{aligned}X_{k+1}&=\Phi_{k+1/k}X_k+\Gamma_kW_k-\Gamma_kS_kR_k^{-1}V_k+\Gamma_kS_kR_k^{-1}V_k\\&=\Phi_{k+1/k}X_k+\Gamma_k\overline{W}_k+\Gamma_kS_kR_k^{-1}V_k\end{aligned} \tag{7-1-46}$$

把$V_k=Z_k-H_kX_k$代入上式，得

$$X_{k+1}=(\Phi_{k+1/k}-\Gamma_kS_kR_k^{-1}H_k)X_k+\Gamma_k\overline{W}_k+\Gamma_kS_kR_k^{-1}Z_k \tag{7-1-47}$$

因为测量值Z_k是已知的，因此可以把$\Gamma_kS_kR_k^{-1}Z_k$看成是已知的控制输入，把$\Gamma_k\overline{W}$看成是输入噪声，于是我们得到等价的数学模型为

$$X_{k+1}=(\Phi_{k+1/k}-\Gamma_kS_kR_k^{-1}H_k)X_k+\Gamma_kS_kR_k^{-1}Z_k+\Gamma_k\overline{W}_k \tag{7-1-48}$$

$$Z_{k+1}=H_{k+1}X_{k+1}+V_{k+1} \tag{7-1-49}$$

对照(7-1-36)~(7-1-40)式，便得到相应的卡尔曼滤波器。

预报估计

$$\hat{X}_{k+1/k}=(\Phi_{k+1/k}-\Gamma_k S_k R_k^{-1}H_k)\hat{X}_k+\Gamma_k S_k R_k^{-1}Z_k$$
$$=\Phi_{k+1/k}\hat{X}_k+\Gamma_k S_k R_k^{-1}(Z_k-H_k\hat{X}_k) \tag{7-1-50}$$

预报误差协方差阵

$$P_{k+1/k}=(\Phi_{k+1/k}-\Gamma_k S_k R_k^{-1}H_k)P_k(\Phi_{k+1/k}-\Gamma_k S_k R_k^{-1}H_k)^T$$
$$+\Gamma_k(Q_k-S_k R_k^{-1}S_k^T)\Gamma_k^T \tag{7-1-51}$$

最佳增益

$$K_{k+1}=P_{k+1/k}H_{k+1}^T[H_{k+1}P_{k+1/k}H_{k+1}^T+R_{k+1}]^{-1} \tag{7-1-52}$$

滤波估计

$$\hat{X}_{k+1}=\hat{X}_{k+1/k}+K_{k+1}(Z_{k+1}-H_{k+1}\hat{X}_{k+1/k}) \tag{7-1-53}$$

滤波误差协方差阵

$$P_{k+1}=(I-K_{k+1}H_{k+1})P_{k+1/k} \tag{7-1-54}$$

我们注意到，在$\hat{X}_{k+1/k}$中包含了k时刻的测量值Z_k，表明在$S_k\neq0$时，在Z_k中包含着有关状态X_{k+1}的信息，这个信息应当加以利用。

§7-2　高斯分布条件下的卡尔曼滤波

在上一节中没有规定各随机变量的概率分布形式，但规定了滤波估计$\hat{X}_k$是预报估计$\hat{X}_{k/k-1}$和测量Z_k的线性组合，得到的估计是线性无偏最小方差估计。现在放弃线性假设，但规定集合$\{X_k\}$和$\{Z_k\}$是联合高斯分布的，并把贝叶斯估计方法应用于过程估计中来，现在将证明，在高斯分布条件下，卡尔曼滤波是最小方差估计，是一切估计中最好的估计。

用符号Z_1^k表示集合$\{Z_k\}$，根据贝叶斯估计中最小方差估计的定义，预报$\hat{X}_{k/k-1}$应当是条件密度函数$p(X_k/Z_1^{k-1})$的均值，即有

$$\hat{X}_{k/k-1}=E[X_k/Z_1^{k-1}]$$
$$=E[(\Phi_{k/k-1}X_{k-1}+W_{k-1})/Z_1^{k-1}] \tag{7-2-1}$$

而滤波估计应当是条件密度$p(X_k/Z_1^k)$的均值。

第一步，假定$p(X_{k-1}/Z_1^{k-1})$是高斯的，现在将证明$p(X_k/Z_1^{k-1})$和$p(X_k/Z_1^k)$也都是高斯的。

设已知

$$p(X_{k-1}/Z_1^{k-1})=\frac{1}{(2\pi)^{\frac{N}{2}}|P_{k-1}|^{\frac{1}{2}}}$$
$$\times\exp\left\{-\frac{1}{2}[X_{k-1}-\hat{X}_{k-1}]^T P_{k-1}^{-1}[X_{k-1}-\hat{X}_{k-1}]\right\} \tag{7-2-2}$$

其中　$\hat{X}_{k-1}=E[X_{k-1}/Z_1^{k-1}]$ (7-2-3)

$$P_{k-1}=E[(X_{k-1}-\hat{X}_{k-1})(X_{k-1}-\hat{X}_{k-1})^T] \tag{7-2-4}$$

已知 $X_k=\Phi_{k/k-1}X_{k-1}+W_{k-1}$，由于 X_{k-1} 和 W_{k-1} 都是高斯的，故而 $p(X_k/Z^*_{k-1})$ 也是高斯的。

根据贝叶斯规则，得

$$p(X_{k-1},\ W_{k-1}/Z^*_{k-1})=\frac{p(X_{k-1},\ W_{k-1},\ Z^*_{k-1})}{p(Z^*_{k-1})} \tag{7-2-5}$$

因为 W_{k-1} 与 X_{k-1}、Z_{k-1} 独立，故上式可写成

$$p(X_{k-1},\ W_{k-1}/Z^*_{k-1})=\frac{p(X_{k-1},\ Z^*_{k-1})p(W_{k-1})}{p(Z^*_k)}$$

$$=p(X_{k-1}/Z^*_{k-1})p(W_{k-1}) \tag{7-2-6}$$

因为已经假定 $p(X_{k-1}/Z^*_{k-1})$ 是高斯的，$p(W_{k-1})$ 是高斯的，故 $p(X_{k-1},\ W_{k-1}/Z^*_{k-1})$ 也是高斯的，从而得知 $p(X_k/Z^*_{k-1})$ 是高斯的。$p(X_k/Z^*_{k-1})$ 的均值为

$$E[X_k/Z^*_{k-1}]=\Phi_{k/k-1}E[X_{k-1}/Z^*_{k-1}]+E[W_{k-1}/Z^*_{k-1}] \tag{7-2-7}$$

但是，W_{k-1} 与 Z^*_k 是不相关的，其条件均值等于无条件均值 $E[W_{k-1}]=0$，于是得

$$E[X_k/Z^*_{k-1}]=\Phi_{k/k-1}E[X_{k-1}/Z^*_{k-1}] \tag{7-2-8}$$

记
$$\hat{X}_{k/k-1}=E[X_k/Z^*_{k-1}] \tag{7-2-9}$$

则有
$$\hat{X}_{k/k-1}=\Phi_{k/k-1}\hat{X}_{k-1} \tag{7-2-10}$$

再令
$$P_{k/k-1}=E[(X_k-\hat{X}_{k/k-1})(X_k-\hat{X}_{k/k-1})^T]$$

容易求得
$$P_{k/k-1}=\Phi_{k/k-1}P_{k-1}\Phi^T_{k/k-1}+Q_{k-1} \tag{7-2-11}$$

于是我们可以写出条件密度函数

$$p(X_k/Z^*_{k-1})=\frac{1}{(2\pi)^{\frac{N}{2}}|P_{k/k-1}|^{\frac{1}{2}}}$$

$$\times\exp\left\{-\frac{1}{2}(X_k-\hat{X}_{k/k-1})^T P^{-1}_{k/k-1}(X_k-\hat{X}_{k/k-1})\right\} \tag{7-2-12}$$

第二步，再考虑得到测量值 Z_k 之后的条件密度函数 $P(X_k/Z^*_k)$。根据贝叶斯规则有

$$p(X_k/Z^*_k)=\frac{p(X_k,\ Z^*_k)}{p(Z^*_k)}$$

$$=\frac{p(X_k,\ Z_k,\ Z^*_{k-1})}{p(Z_k,\ Z^*_{k-1})}$$

$$=\frac{p(Z_k/X_k,\ Z^*_{k-1})\cdot p(X_k,\ Z^*_{k-1})}{p(Z_k/Z^*_{k-1})p(Z^*_{k-1})}$$

$$=\frac{p(Z_k/X_k,\ Z^*_{k-1})\cdot p(X_k/Z^*_{k-1})p(Z^*_{k-1})}{p(Z_k/Z^*_{k-1})p(Z^*_{k-1})}$$

$$=\frac{p(Z_k/X_k,\ Z^*_{k-1})\cdot p(X_k/Z^*_{k-1})}{p(Z_k/Z^*_{k-1})} \tag{7-2-13}$$

现在来求(7-2-13)右边各密度函数的表达式，其中$p(X_k/Z^*_{k-1})$就是(7-2-12)式。

再求$p(Z_k/X_k,\ Z^*_{k-1})$，因为$Z_k=H_kX_k+V_k$，其中V_k是与X_k及Z^*_{k-1}不相关的，且V_k是高斯的，已知$E[V_k]=0$，$E[V_kV_k^T]=R_k$，X_k也是高斯的，故Z_k是高斯的，其条件均值为

$$E[Z_k/X_k,\ Z^*_{k-1}]=H_kX_k \tag{7-2-14}$$

得出条件密度函数为

$$p(Z_k/X_k,\ Z^*_{k-1})=\frac{1}{(2\pi)^{\frac{m}{2}}|R_k|^{\frac{1}{2}}}$$

$$\times\exp\left\{-\frac{1}{2}(Z_k-H_kX_k)^TR_k^{-1}(Z_k-H_kX_k)\right\} \tag{7-2-15}$$

再计算分母$p(Z_k/Z^*_{k-1})$，因为

$$p(X_kV_k/Z^*_{k-1})=p(V_k/X_kZ^*_{k-1})\cdot p(X_k/Z^*_{k-1}) \tag{7-2-16}$$

但V_k与X_k及Z^*_{k-1}独立，故

$$p(X_k,\ V_k/Z^*_{k-1})=p(V_k)p(X_k/Z^*_{k-1}) \tag{7-2-17}$$

因为(7-2-17)右边两项都是高斯的，其积也是高斯的，故$p(X_k,\ V_k/Z^*_{k-1})$是高斯的，其均值为

$$E[Z_k/Z^*_{k-1}]=E[H_kX_k+V_k/Z^*_{k-1}]$$

$$=H_kE[X_k/Z^*_{k-1}]=H_k\hat{X}_{k/k-1} \tag{7-2-18}$$

其协方差阵为

$$E[(Z_k-H_k\hat{X}_{k/k-1})(Z_k-H_k\hat{X}_{k/k-1})^T]$$

$$=H_kP_{k/k-1}H_k^T+R_k \tag{7-2-19}$$

于是得出密度函数

$$p(Z_k/Z^*_{k-1})=\frac{1}{(2\pi)^{\frac{m}{2}}|H_kP_{k/k-1}H_k^T+R_k|^{\frac{1}{2}}} \tag{7-2-20}$$

$$\times\exp\left\{-\frac{1}{2}(Z_k-H_k\hat{X}_{k/k-1})^T(H_kP_{k/k-1}H_k^T+R_k)^{-1}(Z_k-H_k\hat{X}_{k/k-1})\right\}$$

把相应的各密度函数代入(7-2-13)式中，得

$$p(X_k/Z^*_k)=\frac{|H_kP_{k/k-1}H_k^T+R_k|^{\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}|P_{k/k-1}|^{\frac{1}{2}}|R_k|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\cdot)\right\} \tag{7-2-21}$$

其中 $(\cdot)=(Z_k-H_kX_k)^TR_k^{-1}(Z_k-H_kX_k)$

$$+(X_k-\hat{X}_{k/k-1})^TP_{k/k-1}^{-1}(X_k-\hat{X}_{k/k-1})$$

$$-(Z_k-H_k\hat{X}_{k/k-1})^T(H_kP_{k/k-1}H_k^T+R_k)^{-1}(Z_k-H_k\hat{X}_{k/k-1}) \tag{7-2-22}$$

现在对(7-2-22)式加以整理，把三个二次型合并成一个二次型，这里要用到附录D中三个矩阵等式

$$(P^{-1}+H^TR^{-1}H)^{-1}=P-PH^T(HPH^T+R)^{-1}HP$$

$$=(1-PH^TR^{-1}H)P \tag{7-2-23}$$

$$(P^{-1}+H^TRH)^{-1}H^TR^{-1}=PH^T(HPH^T+R)^{-1} \tag{7-2-24}$$

$$H(P^{-1}+H^TR^{-1}H)H^T = R - R(HPH^T+R)^{-1}R \tag{7-2-25}$$

略去时间下标k，记$\hat{X}_{k/k-1}=\hat{X}_1$，$P_{k/k-1}=P_1$，

则
$$(\cdot) = X^T(P_1^{-1}+H^TR^{-1}H)X - 2X^T(P_1^{-1}\hat{X}_1+H^TR^{-1}Z)$$
$$+Z^T[R^{-1}-(HP_1H^T+R)^{-1}]Z+2\hat{X}_1^TH^T(HP_1H^T+R)^{-1}Z$$
$$+\hat{X}_1^T[P_1^{-1}+H^T(HP_1H^T+R)^{-1}H]\hat{X}_1 \tag{7-2-26}$$

考察(7-2-26)式右边第三项，利用等式(7-2-25)得

$$R^{-1}-(HP_1H^T+R)^{-1} = R^{-1}RR^{-1}-R^{-1}R(HRH^T+R)RR^{-1}$$
$$=R^{-1}H(P_1^{-1}+H^TR^{-1}H)^{-1}R^{-1} \tag{7-2-27}$$

故第三项可写成

$$Z^TR^{-1}H(P_1^{-1}+H^TR^{-1}H)^{-1}H^TR^{-1}Z \tag{7-2-28}$$

考察(7-2-26)式右边第四项，利用等式(7-2-24)式，得

$$P_1^{-1}(P_1^{-1}+H^TR^{-1}H)^{-1}HR^{-1}=P_1^{-1}P_1H^T(HP_1H^T+R)^{-1}$$
$$=H^T(HP_1H^T+R)^{-1} \tag{7-2-29}$$

故第四项可写成

$$2\hat{X}_1^TP_1^{-1}(P_1^{-1}+H^TR^{-1}H)^{-1}H^TR^{-1}Z \tag{7-2-30}$$

再考察(7-2-26)右边第五项，利用等式(7-2-23)，得

$$P_1^{-1}(P_1^{-1}+H^TR^{-1}H)^{-1}P_1^{-1}=P_1^{-1}-H^T(HP_1H^T+R)^{-1}H \tag{7-2-31}$$

故第五项可写成

$$\hat{X}_1^TP_1^{-1}(P_1^{-1}+H^TR^{-1}H)^{-1}P_1^{-1}\hat{X}_1 \tag{7-2-32}$$

把第三、四、五项合并，再加上第一、二项，得

$$(\cdot)=X^T(P_1^{-1}+H^TR^{-1}H)X-2X^T(P_1^{-1}\hat{X}_1+H^TR^{-1}Z)$$
$$+(Z^TR^{-1}H+\hat{X}_1^TP_1^{-1})(P_1^{-1}+H^TR^{-1}H)^{-1}(H^TR^{-1}Z+P_1^{-1}\hat{X}_1) \tag{7-2-33}$$

令
$$A=P_1^{-1}+H^TR^{-1}H \tag{7-2-34}$$
$$a=P_1^{-1}\hat{X}_1+H^TR^{-1}Z \tag{7-2-35}$$

则
$$(\cdot)=X^TAX-2X^Ta+a^TA^{-1}a$$
$$=X^TAX-2X^TAA^{-1}a+a^TA^{-1}AA^{-1}a$$
$$=(X-A^{-1}a)^T(X-A^{-1}a) \tag{7-2-36}$$

于是，我们得到$p(X_k/Z_1^k)$的条件均值为

$$\hat{X}_k=E[X_k/Z_1^k]=A^{-1}a$$
$$=(P_{k/k-1}^{-1}+H_k^TR_k^{-1}H_k)^{-1}(P_{k/k-1}\hat{X}_{k/k-1}+H_k^TR_k^{-1}H_k) \tag{7-2-37}$$

$p(X_k/Z_1^k)$的协方差阵为

$$P_k=A^{-1}=(P_{k/k-1}^{-1}+H_k^TR_k^{-1}H_k)^{-1} \tag{7-2-38}$$

从以上结果可知，若已知$p(X_{k-1}/Z_1^{k-1})$是高斯的，其条件均值为$\hat{X}_{k-1}$，条件方差阵为P_{k-1}，那么经过时间传递并加上新测量值之后，密度函数$p(X_k/Z_1^k)$也是高斯的，其条件均值$X_k=E[X_k/Z_1^k]$如(7-2-33)式所示，其协方差阵P_k如(7-2-38)式所示。

要求初始条件X_0也是高斯的，其均值和方差应是已知的。

$$\hat{X}_0 = E[X_0] = \overline{X}_0 \tag{7-2-39}$$

$$P_0 = E[(X_0 - \overline{X}_0)(X_0 - \overline{X}_0)^T] = \text{var}\, X_0 \tag{7-2-40}$$

根据最小方差估计的定义，后验密度的条件均值就是最小方差估计，所以$\hat{X}_k = E[X_k/Z_k^*]$就是最小方差估计。

我们注意到，(7-2-37)式所示之$\hat{X}_k$正好就是卡尔曼滤波估计。由此可以得出结论：在高斯分布条件下，卡尔曼滤波是最小方差估计，注意这里取消了"线性"二字的限制，所以它是一切估计中最好的估计。

由于$p(X_k/Z_k^*)$是高斯的，故最小方差估计同时又是最大后验估计。

§7-3 卡尔曼滤波与线性最小方差估计

重写第三章中线性最小方差估计表达式

$$\hat{X}_{lmv} = EX + \text{cov}(X,\ Z)(\text{var}Z)^{-1}(Z - EZ)$$

推广至时间序列估计时，加上时标k，得

$$\hat{X}_k = EX_k + \text{cov}(X_k,\ Z_k)(\text{var}Z_k)^{-1}(Z_k - EZ_k) \tag{7-3-1}$$

其中EX_k、EZ_k为先验均值。若已得到$k-1$步的无偏估计$\hat{X}_{k-1}$，则k步的先验均值为

$$EX_k = \Phi_{k,k-1}\hat{X}_{k-1} = \hat{X}_{k/k-1} \tag{7-3-2}$$

$$EZ_k = H_k EX_k = H_k \hat{X}_{k/k-1} \tag{7-3-3}$$

又

$$\begin{aligned} \text{cov}(X_k,\ Z_k) &= E[(X_k - EX_k)(Z_k - EZ_k)^T] \\ &= E[(X_k - \hat{X}_{k/k-1})(H_k X_k + V_k - H_k \hat{X}_{k/k-1})^T] \\ &= E[(X_k - \hat{X}_{k/k-1})(X_k - \hat{X}_{k/k-1})^T H_k^T] \\ &\quad + E[(X_k - \hat{X}_{k/k-1})V_k^T] \end{aligned}$$

上式右边第二项等于零，故有

$$\text{cov}(X_k,\ Z_k) = P_{k/k-1}H_k^T \tag{7-3-4}$$

又

$$\begin{aligned} \text{var}Z_k &= E[(Z_k - EZ_k)(Z_k - EZ_k)^T] \\ &= E[(Z_k - H_k\hat{X}_{k/k-1})(Z_k - H_k\hat{X}_{k/k-1})^T] \end{aligned}$$

代入关系式$Z_k = H_k X_k + V_k$得

$$\text{var}Z_k = H_k P_{k/k-1} H_k^T + R_k \tag{7-3-5}$$

把以上结果代入(7-3-1)式中，得

$$\hat{X}_k = \hat{X}_{k/k-1} + P_{k/k-1}H_k^T[H_k P_{k/k-1} H_k^T + R_k]^{-1}[Z_k - H_k\hat{X}_{k/k-1}] \tag{7-3-6}$$

显见上式即是卡尔曼滤波估计表达式。

再重写方差P_{lmv}表达式

$$P_{lmv} = \text{var}\, X - \text{cov}(X,\ Z)(\text{var}Z)^{-1}\text{cov}(Z,\ X)$$

加上时标k，得

$$P_k = \text{var}\, X_k - \text{cov}(X_k,\ Z_k)(\text{var}Z_k)^{-1}\text{cov}(Z_k,\ X_k) \tag{7-3-7}$$

其中 $\operatorname{var} X_k = E[(X_k - \hat{X}_{k/k-1})(X_k - X_{k/k-1})^T] = P_{k/k-1}$

$\operatorname{cov}(X_k, Z_k) = P_{k/k-1} H_k^T$

$\operatorname{cov}(Z_k, X_k) = H_k P_{k/k-1}$

代入(7-3-7)式，得

$$P_k = P_{k/k-1} - P_{k/k-1} H_k^T [H_k P_{k/k-1} H_k^T + R_k]^{-1} H_k P_{k/k-1} \tag{7-3-8}$$

已知滤波器增益为

$$K_k = P_{k/k-1} H_k^T [H_k P_{k/k-1} H_k^T + R_k]^{-1} \tag{7-3-9}$$

故有

$$\hat{X}_k = \hat{X}_{k/k-1} + K_k [Z_k - H_k \hat{X}_{k/k-1}] \tag{7-3-10}$$

$$P_k = [1 - K_k H_k] P_{k/k-1} \tag{7-3-11}$$

由以上讨论可见，卡尔曼滤波是第三章中线性最小方差估计在过程估计中的推广。

§7-4 卡尔曼滤波与最优加权最小二乘估计

在第二章中研究了对测量集合的最优加权最小二乘估计，现在来研究对测量序列的最优加权最小二乘估计。

设信号状态模型为

$$X_k = \Phi_{k/k-1} X_{k-1} \tag{7-4-1}$$

为方便起见，这里暂且假设无状态噪声。

测量是有噪声的，测量序列为

$$Z_k = H_k X_k + V_k \tag{7-4-2}$$

已知V_k为零均值白噪声，方差为$E[V_k, V_k^T] = R_k$。

已知$k=0$时的初始估计为$\hat{X}_0$，其方差为P_0，现在求状态X_k的最优加权最小二乘估计$\hat{X}_k$。

各个时刻估计误差的最优加权平方和为

$$J = (\hat{X}_0 - \Phi_{0/k} \hat{X}_k)^T P_0^{-1} (\hat{X}_0 - \Phi_{0/k} \hat{X}_k) + \sum_{i=1}^{k} (Z_i - H_i \Phi_{i/k} \hat{X}_k)^T R_i^{-1} (Z_i - H_i \Phi_{i/k} \hat{X}_k) \tag{7-4-3}$$

其中 $\Phi_{i/k} = \Phi_{i/i+1} \cdot \Phi_{i+1/i+2} \cdots \Phi_{k-1/k}$

$\Phi_{i/k}^{-1} = \Phi_{k/i}$

求使J达到极小时的$\hat{X}_k$，由

$$\frac{\partial J}{\partial \hat{X}_k} = 0 \tag{7-4-4}$$

可求出

$$\hat{X}_k = \left[\Phi_{0/k}^T P_0^{-1} \Phi_{0/k} + \sum_{i=1}^{k} \Phi_{i/k}^T H_i^T R_i^{-1} H_i \Phi_{i/k}\right]^{-1} \times \left[\Phi_{0/k}^T P_0^{-1} \hat{X}_0 + \sum_{i=1}^{k} \Phi_{i/k}^T H_i^T R_i^{-1} Z_i\right] \tag{7-4-5}$$

记
$$\hat{X}_{k/0}=\Phi_{k/0}\hat{X}_0$$
$$P_{k/0}=\Phi_{k/0}P_0\Phi_{k/0}^T$$
则有
$$\hat{X}=\left[P_{k/0}^{-1}+\sum_{i=1}^{k}\Phi_{i/k}^T H_i^T R_i^{-1} H_i \Phi_{i/k}\right]^{-1}$$
$$\times\left[P_{k/0}^{-1}\hat{X}_{k/0}+\sum_{i=1}^{k}\Phi_{i/k}^T H_i^T R_i^{-1} Z_i\right] \tag{7-4-6}$$
估计误差方差为
$$P_k=\left[P_{k/0}^{-1}+\sum_{i=1}^{k}\Phi_{i/k}^T H_i^T R_i^{-1} H_i \Phi_{i/k}\right]^{-1} \tag{7-4-7}$$

我们注意到，若$\Phi_{k/k-1}=1$，$H_k=h_i$则上述成果就是第二章中的最优加权最小二乘估计表达式。

若已知t_{k-1}时刻的估计$\hat{X}_{k-1}$和P_{k-1}，在加入新测量Z_k之后，就可以得到t_k时刻的估计$\hat{X}_k$、P_k。对(7-4-6)、(7-4-7)式化简，取$\hat{X}_{k-1}$、P_{k-1}作为初值，于是有
$$\hat{X}_k=[P_{k/k-1}^{-1}+H_k^T R_k^{-1} H_k]^{-1}[P_{k/k-1}^{-1}\hat{X}_{k/k-1}+H_k^T R_k^{-1} Z_k] \tag{7-4-8}$$
$$P_k=[P_{k/k-1}^{-1}+H_k^T R_k^{-1} H_k]^{-1} \tag{7-4-9}$$
利用附录D的矩阵恒等式，(7-4-9)式改写成
$$P_k=P_{k/k-1}-P_{k/k-1}H_k^T[H_k P_{k/k-1} H_k^T+R_k]^{-1}H_k P_{k/k-1} \tag{7-4-10}$$
记
$$K_k=P_{k/k-1}H_k^T[H_k P_{k/k-1}H_k^T+R_k]^{-1} \tag{7-4-11}$$
代入(7-4-10)式中，得
$$P_k=(1-K_k H_k)P_{k/k-1} \tag{7-4-12}$$
其中
$$P_{k/k-1}=\Phi_{k/k-1}P_{k-1}\Phi_{k/k-1}^T \tag{7-4-13}$$
再由附录D中矩阵等式可证明
$$K_k=P_{k/k-1}H_k^T[H_k P_{k/k-1}H_k^T+R_k]^{-1}$$
$$=[P_{k/k-1}^{-1}+H_k^T R_k^{-1} H_k]^{-1}H_k^T R_k^{-1} \tag{7-4-14}$$
故有
$$K_k=P_k H_k^T R_k^{-1} \tag{7-4-15}$$
记
$$\hat{X}_{k/k-1}=\Phi_{k/k-1}\hat{X}_{k-1} \tag{7-4-16}$$
同理可得
$$\hat{X}_k=(1-K_k H_k)\hat{X}_{k/k-1}+K_k Z_k \tag{7-4-17}$$
总结一下，一步递推最优加权最小二乘估计公式为
$$\left.\begin{aligned}&\hat{X}_{k/k-1}=\Phi_{k/k-1}\hat{X}_{k}\\&P_{k/k-1}=\Phi_{k/k-1}P_{k-1}\Phi_{k/k-1}^T\\&K_k=P_{k/k-1}H_k^T[H_k P_{k/k-1}H_k^T+R_k]^{-1}\end{aligned}\right\} \tag{7-4-18}$$

$$\left.\begin{aligned}\hat{X}_k &= (1-K_kH_k)\hat{X}_{k/k-1}+K_kZ_k\\ P_k &= (1-K_kH_k)P_{k/k-1}\end{aligned}\right\}\qquad(7\text{-}4\text{-}18)$$

对于有噪声的状态方程

$$X_k=\Phi_{k/k-1}X_{k-1}+W_{k-1}\qquad(7\text{-}4\text{-}19)$$

其中W_k为零均值白噪声，方差为Q_k，在(7-4-19)式中的$P_{k/k-1}$应修改为

$$P_{k/k-1}=\Phi_{k/k-1}P_{k-1}\Phi^T_{k/k-1}+Q_{k-1}\qquad(7\text{-}4\text{-}20)$$

由以上分析可知，卡尔曼滤波与时间序列的最优加权最小二乘估计是等价的。

对于无状态噪声的等速度信号模型

$$\left.\begin{aligned}\begin{bmatrix}x_k\\ \dot{x}_k\end{bmatrix} &= \begin{bmatrix}1 & T\\ 0 & 1\end{bmatrix}\begin{bmatrix}x_{k-1}\\ \dot{x}_{k-1}\end{bmatrix}\\ z_k &= x_k+v_k\end{aligned}\right\}\qquad(7\text{-}4\text{-}21)$$

真实距离x_k、测量z_k、估计距离$\hat{x}_k$的关系如图7-4-1所示。

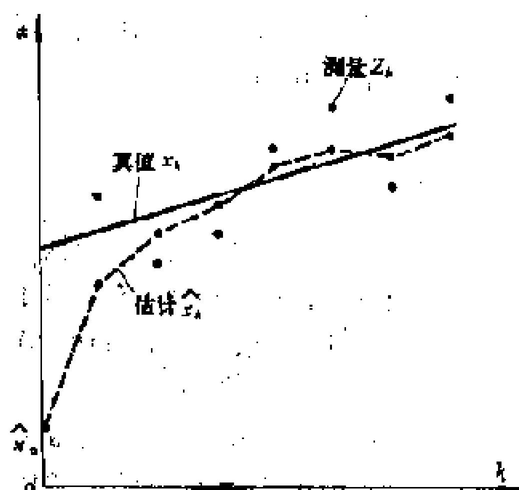


图7-4-1 无状态噪声时的距离估计

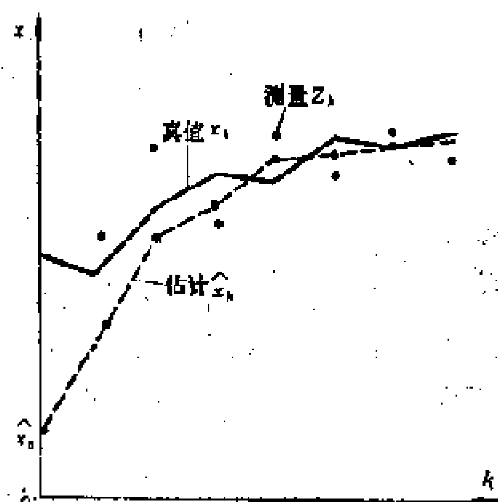


图7-4-2 有状态噪声时的距离估计

对于有状态噪声的信号模型

$$\left.\begin{aligned}\begin{bmatrix}x_k\\ \dot{x}_k\end{bmatrix} &= \begin{bmatrix}1 & T\\ 0 & 1\end{bmatrix}\begin{bmatrix}x_{k-1}\\ \dot{x}_{k-1}\end{bmatrix}+\begin{bmatrix}0\\ w_{k-1}\end{bmatrix}\\ z_k &= x_k+v_k\end{aligned}\right\}\qquad(7\text{-}4\text{-}22)$$

x_k、z_k、$\hat{x}_k$的关系如图7-4-2所示。

从递推最优加权最小二乘估计出发，也可得出卡尔曼滤波。假设已经得到了$k-1$时刻的最优估计$\hat{X}_{k-1}$和P_{k-1}，有了新测量Z_k之后，可以在$\hat{X}_{k-1}$和Z_k的基础上求出$\hat{X}_k$和P_k，不必利用过去的测量。

把Z_k、$\hat{X}_{k/k-1}$看成是对X_k的两个无偏独立观测值，有

$$\left.\begin{aligned}Z_k &= H_kX_k+V_k\\ \hat{X}_{k/k-1} &= X_k+\tilde{X}_{k/k-1}\end{aligned}\right\}\qquad(7\text{-}4\text{-}23)$$

其中 $\hat{X}_{k/k-1}=\Phi_{k/k-1}\hat{X}_{k-1}$，$\tilde{X}_{k/k-1}$为预报误差。

已知
$$E[V_k V_k^T]=R_k$$
$$E[\tilde{X}_{k/k-1}\tilde{X}_{k/k-1}^T]=\Phi_{k/k-1}P_{k-1}\Phi_{k/k-1}^T+Q_{k-1}=P_{k/k-1}$$

现在来求 X_k 的最优加权最小二乘估计$\hat{X}_k$。

记
$$\mathrm{Z}=\begin{bmatrix} Z_k \\ \hat{X}_{k/k-1} \end{bmatrix},\quad \mathrm{H}=\begin{bmatrix} H_k \\ I \end{bmatrix},\quad \mathrm{R}=\begin{bmatrix} R_k & 0 \\ 0 & P_{k/k-1} \end{bmatrix} \tag{7-4-24}$$

则有
$$\hat{X}_k=[\mathrm{H}^T\mathrm{R}^{-1}\mathrm{H}]^{-1}\mathrm{H}^T\mathrm{R}^{-1}\mathrm{Z} \tag{7-4-25}$$
$$P_k=[\mathrm{H}^T\mathrm{R}^{-1}\mathrm{H}]^{-1} \tag{7-4-26}$$

把以上各表达式代入，得
$$\hat{X}_k=[H_k^T R_k^{-1}H_k+P_{k/k-1}^{-1}]^{-1}[H_k^T R_k^{-1}Z_k+P_{k/k-1}^{-1}\hat{X}_{k/k-1}] \tag{7-4-27}$$
$$P_k=[H_k^T R_k^{-1}H_k+P_{k/k-1}^{-1}] \tag{7-4-28}$$

再应用矩阵恒等式(附录D)，就可以得到卡尔曼滤波公式。

由以上分析可知，线性信号模型的卡尔曼滤波实际上与最优加权最小二乘估计是等价的。因此，卡尔曼滤波解的存在条件就是最优加权最小二乘估计解的存在条件。

§7-5 卡尔曼滤波解的存在条件

卡尔曼滤波解是从信号模型按照线性无偏最小方差准则导出的，现在需要研究两个问题：1.是不是任意线性模型都存在卡尔曼滤波解，或者说什么样的信号模型才存在卡尔曼滤波解。2.如果卡尔曼滤波解存在，表明卡尔曼滤波器是一稳定反馈系统，它有什么样的稳定性性质。本节先研究第一个问题。

直接从卡尔曼滤波公式求解的存在条件有很大困难。根据上节中卡尔曼滤波与最优加权最小二乘估计等价性的结论，求出线性信号模型的最优加权最小二乘估计的存在条件，也就得出了卡尔曼滤波解的存在条件。

列出各测量值与状态 X_k之间的关系

$$
\begin{aligned}
Z_k &= H_k X_k+V_k \\
Z_{k-1} &= H_{k-1}\Phi_{k-1/k}X_k-H_{k-1}\Phi_{k-1/k}\Gamma_{k-1}W_{k-1}+V_{k-1} \\
Z_{k-2} &= H_{k-2}\Phi_{k-2/k}X_k \\
&\quad -H_{k-2}\Phi_{k-2/k}(\Phi_{k/k-1}\Gamma_{k-2}W_{k-2}+\Gamma_{k-1}W_{k-1})+V_{k-2} \\
&\cdots\cdots \\
Z_{k-i} &= H_{k-i}\Phi_{k-i/k}X_k-H_{k-i}\Phi_{k-i/k}\left(\sum_{l=1}^{i}\Phi_{k/k-l+1}\Gamma_{k-l}W_{k-l}\right)+V_{k-i} \\
&\cdots\cdots \\
Z_1 &= H_1\Phi_{1/k}X_k-H_1\Phi_{1/k}\left(\sum_{l=1}^{k-1}\Phi_{k/k-l+1}\Gamma_{k-l}W_{k-l}\right)+V_1 \\
Z_0 &= H_0 X_0+H_0\tilde{X}_0=H_0\hat{X}_0 \\
&= H_0\Phi_{0/k}X_k-H_0\Phi_{0/k}\left(\sum_{l=1}^{k}\Phi_{k/k-l+1}\Gamma_{k-l}W_{k-l}\right)+H_0\tilde{X}_0
\end{aligned}
\tag{7-5-1}
$$

记

$$
H=\begin{bmatrix} H_k \\ H_{k-1}\Phi_{k-1/k} \\ \vdots \\ H_1\Phi_{1/k} \\ H_0\Phi_{0/k} \end{bmatrix},\quad V=\begin{bmatrix} V_k \\ V_{k-1} \\ \vdots \\ V_1 \\ H_0\widetilde{X}_0 \end{bmatrix}
$$

$$
W=\begin{bmatrix} 0 \\ H_{k-1}\Phi_{k-1/k}\sum_{l=1}^{1}\Phi_{k/k-l+1}\Gamma_{k-l}W_{k-l} \\ H_{k-2}\Phi_{k-2/k}\sum_{l=1}^{2}\Phi_{k/k-l+1}\Gamma_{k-l}W_{k-l} \\ \vdots \\ H_1\Phi_{1/k}\sum_{l=1}^{k-1}\Phi_{k/k-l+1}\Gamma_{k-l}W_{k-l} \\ H_0\Phi_{0/k}\sum_{l=1}^{k}\Phi_{k/k-l+1}\Gamma_{k-l}W_{k-l} \end{bmatrix},\quad Z=\begin{bmatrix} Z_k \\ Z_{k-1} \\ \vdots \\ Z_1 \\ Z_0 \end{bmatrix} \tag{7-5-2}
$$

于是得到方程式

$$
Z = HZ_k - W + V \tag{7-5-3}
$$

由(7-5-3)式可求得 X_k 的最优加权最小二乘估计。

已知 $E[W]=0$，$E[V]=0$，$E[VW^T]=0$

令

$$
\left.\begin{aligned} S &= E[WW^T] \\ R &= E[VV^T] \\ &= \begin{bmatrix} R_k & & & 0 \\ & R_{k-1} & & \\ & & \ddots\ R_1 & \\ 0 & & & R_0 \end{bmatrix},\quad R_0 = H_0P_0H_0^T \end{aligned}\right\} \tag{7-5-4}
$$

于是有 $E[(V-W)(V-W)^T] = R + S$ (7-5-5)

故得

$$
\hat{X}_k = [H^T(R+S)^{-1}H]^{-1}H^T(R+S)^{-1}Z \tag{7-5-6}
$$

$$
P_k = [H^T(R+S)^{-1}H]^{-1} \tag{7-5-7}
$$

由以上二式可知，存在最优加权最小二乘估计的充分必要条件是：

1. $R=E[VV^T]$ 为正定和有界，这一点通常是满足的。

2. $S=E[WW^T]$ 为正定和有界。

此条件等价于要求矩阵

$$
E\left[\left(\sum_{l=1}^{k}\Phi_{k/k-l+1}\Gamma_{k-l}W_{k-l}\right)\left(\sum_{l=1}^{k}\Phi_{k/k-l+1}\Gamma_{k-l}W_{k-l}\right)^T\right]
$$

为正定和有界，它又等价于要求矩阵

$$
C = \sum_{l=1}^{k}\Phi_{k/k-l+1}\Gamma_{k-l}Q_{k-l}\Gamma_{k-l}^T\Phi_{k/k-l+1}^T \tag{7-5-8}
$$

为正定和有界。这里利用了状态噪声W_k的白噪声性质。

对(7-5-8)式作时标变换。令$j=k-l+1$，得

$$C=\sum_{j=1}^{k}\Phi_{k/j}\Gamma_{j-1}Q_{j-1}\Gamma_{j-1}^{T}\Phi_{k/j}^{T} \tag{7-5-9}$$

矩阵为正定和有界、可逆、满秩，这三种提法是等价的。由第六章知道，若在所有时刻k上，矩阵C的秩等于N（X_k的维数），则称信号模型是完全可控制的。因此，要求矩阵S为正定和有界等价于要求信号模型为完全可控制的。

3. 矩阵$H^T(R+S)^{-1}H$为可逆，即P_k为正定和有界。

不失一般性，可假设状态噪声$W_k=0$，故有$S=0$，得出

$$P_k=(H^TR^{-1}H)^{-1} \tag{7-5-10}$$

要求矩阵$H^TR^{-1}H$为可逆等价于要求矩阵

$$M=\sum_{j=1}^{k}\Phi_{j/k}^{T}H_j^TR_j^{-1}H_j\Phi_{j/k} \tag{7-5-11}$$

为可逆。

由第六章知道，若在所有时刻k上，矩阵M的秩等于N（状态X_k的维数），或可逆则称信号模型是完全可观测的。因此，要求矩阵M为可逆等价于要求信号模型是完全可观测的。

以上分析表明，存在最优加权最小二乘估计的充分必要条件是信号模型为完全可控制的和完全可观测的。再由§7-5知道，卡尔曼滤波解是与最优加权最小二乘估计是等价的，所以信号模型的完全可观测性与完全可控制性保证了卡尔曼滤波解的存在。再由最优加权最小二乘估计的唯一性可知，卡尔曼滤波解也是唯一的。根据上述结论，可得出如下的卡尔曼滤波存在性定理：

定理1 若信号模型是完全可观测的与完全可控制的，那末其卡尔曼滤波器是一稳定反馈系统，其卡尔曼滤波解存在且是唯一的。

定理2 若卡尔曼滤波器是一稳定的反馈系统，其滤波转移矩阵

$$\Psi_{k/k-1}=(I-K_kH_k)\Phi_{k/k-1}$$

的特征值恒位于单位圆内。

证：在卡尔曼滤波器中，已知

$$\begin{aligned}\hat{X}_k&=(I-K_kH_k)\Phi_{k/k-1}\hat{X}_{k-1}+K_kZ_k\\&=\Psi_{k/k-1}\hat{X}_{k-1}+K_kZ_k\end{aligned} \tag{7-5-12}$$

两边取Z变换，得

$$\hat{X}_k(Z)=\Psi_{k/k-1}Z^{-1}\hat{X}_k(Z)+K_kZ_k(Z) \tag{7-5-13}$$

故在k时刻，卡尔曼滤波器的传递函数为

$$H_k(Z)=[I-\Psi_{k/k-1}Z^{-1}]^{-1}K_k \tag{7-5-14}$$

由线性系统理论可知，滤波器成为稳定系统的充分必要条件是$\Psi_{k/k-1}$的全部特征值λ_i均位于单位圆内，即有

$$|\lambda_i(\Psi_{k/k-1})|<1 \tag{7-5-15}$$

对于一个稳定的卡尔曼滤波器，当输入Z_k有界时，其输出$\hat{X}_k$也必是有界的，协方差

阵P_k也必定是正定和有界的，或者说，其卡尔曼滤波解一定存在。

对于定常信号模型而言，由第六章知道，完全可控制性条件为矩阵

$$C=[\Gamma,\ \Phi\Gamma,\ \Phi^2\Gamma,\ \cdots,\ \Phi^{N-1}\Gamma] \tag{7-5-16}$$

的秩等于N(状态X的维数)。完全可观测性条件为矩阵

$$M=\begin{bmatrix} H \\ H\Phi \\ H\Phi^2 \\ \vdots \\ H\Phi^{N-1} \end{bmatrix} \tag{7-5-17}$$

的秩等于N。所以，对于定常信号模型而言，其卡尔曼滤波解存在的充分必要条件是矩阵C和M的秩等于N(或矩阵C与M为可逆)。

§7-6　卡尔曼滤波器的稳定性

上节已得到了卡尔曼滤波器的存在性定理，本节中将研究卡尔曼滤波的稳态性质，就是当滤波步数充分大之后，滤波器有些什么极限性质。

重新写出信号模型及其卡尔曼滤波器如下：

信号模型为

$$X_k=\Phi_{k/k-1}X_{k-1}+W_{k-1}$$
$$Z_k=H_kX_k+V_k$$
$$E[W_k]=0,\ \mathrm{var}[W_k]=Q_k\delta_{kj},\ E[V_k]=0,\ \mathrm{var}[V_k]=R_k\delta_{kj}$$

初值是$\hat{X}_0$、P_0。

相应的卡尔曼滤波器为

$$\hat{X}_{k/k-1}=\Phi_{k/k-1}\hat{X}_{k-1} \tag{7-6-1}$$

$$P_{k/k-1}=\Phi_{k/k-1}P_{k-1}\Phi^T_{k/k-1}+Q_{k-1} \tag{7-6-2}$$

$$\left.\begin{aligned} K_k&=P_{k/k-1}H_k^T[H_kP_{k/k-1}H_k^T+R_k]^{-1}\\ &=[P^{-1}_{k/k-1}+H_k^TR_k^{-1}H_k]^{-1}H_k^TR_k^{-1}\\ &=P_kH_k^TR_k^{-1}\end{aligned}\right\} \tag{7-6-3}$$

$$\left.\begin{aligned} \hat{X}_k&=\hat{X}_{k/k-1}+K_k[Z_k-H_k\hat{X}_{k/k-1}]\\ &=[1-K_kH_k]\hat{X}_{k/k-1}+K_kZ_k\\ &=P_k^{-1}[P^{-1}_{k/k-1}\hat{X}_{k/k-1}+H_k^TR_k^{-1}H_k]\end{aligned}\right\} \tag{7-6-4}$$

$$\left.\begin{aligned} P_k&=[P^{-1}_{k/k-1}+H_k^TR_k^{-1}H_k]^{-1}\\ &=(1-K_kH_k)P_{k/k-1}(1-K_kH_k)^T+K_kR_kK_k^T\\ &=(1-K_kH_k)P_{k/k-1}\end{aligned}\right\} \tag{7-6-5}$$

假设上述卡尔曼滤波器是满足上节中的存在性定理的，那么它具有以下的稳态性质：

定理3 当滤波步数充分大之后，$\hat{X}_k$将不依赖于初值$\hat{X}_0$的选取。

证：因为

$$\hat{X}_k = \Psi_{k/k-1}\hat{X}_{k-1} + K_k Z_k$$
$$= \Psi_{k/k-1}\Psi_{k-1/k-2}\hat{X}_{k-2} + \Psi_{k/k-1}K_{k-1}Z_{k-1} + K_k Z_k$$
$$= \Psi_{k/k-1}\cdots\Psi_{1/0}\hat{X}_0 + K_k Z_k + \Psi_{k/k-1}K_{k-1}Z_{k-1} + \cdots \qquad (7\text{-}6\text{-}6)$$

由定理2可知，当时间$k\to\infty$时，上式中第一项趋于零，即初值$\hat{X}_0$的影响会消失，$\hat{X}_k$将不依赖于$\hat{X}_0$值的大小。此性质对于时变或定常模型均适用。

定理4 对于同一滤波器，从不同初值$P_0(1)$和$P_0(2)$出发，若$P_0(1)>P_0(2)$，则恒有

$$P_k(1) > P_k(2) \qquad (7\text{-}6\text{-}7)$$

证：已知

$$P_k(1) = [H_k^T R_k^{-1} H_k + P_{k/k-1}^{-1}(1)]^{-1} \qquad (7\text{-}6\text{-}8)$$
$$P_k(2) = [H_k^T R_k^{-1} H_k + P_{k/k-1}^{-1}(2)]^{-1} \qquad (7\text{-}6\text{-}9)$$

若 $P_{k/k-1}(1) > P_{k/k-1}(2)$

则有 $P_{k/k-1}^{-1}(1) < P_{k/k-1}^{-1}(2)$ 和 $P_k(1) > P_k(2)$

今已知 $P_0(1)>P_0(2)$，则必有$P_1(1)>P_1(2)$，故从$P_0(1)$、$P_0(2)$，出发，可以递推算出$P_k(1)$和$P_k(2)$，必定满足$P_k(1)>P_k(2)$。此性质对时变或定常模型同样适用。

协方差矩阵的大小可以用它的迹表示，例如，若$\mathrm{tr}P_0(1)>\mathrm{tr}P_0(2)$，则可以说成$P_0(1)>P_0(2)$。

定理5 在滤波步数充分大之后，滤波误差方差阵P_k将不依赖于初值P_0的选取。

证：由已知公式

$$\left.\begin{aligned} P_k &= [H_k^T R_k^{-1} H_k + P_{k/k-1}^{-1}]^{-1} \\ P_{k/k-1} &= \Phi_{k/k-1}P_{k-1}\Phi_{k/k-1}^T + Q_{k-1} \end{aligned}\right\} \qquad (7\text{-}6\text{-}10)$$

从P_0出发，可推得

$$P_{1/0}^{-1} = [\Phi_{1/0}P_0\Phi_{1/0}^T + Q_0]^{-1}$$
$$P_1 = [H_1^T R_1^{-1} H_1 + P_{1/0}^{-1}]^{-1}$$
$$= [H_1^T R_1^{-1} H_1 + (\Phi_{1/0}P_0\Phi_{1/0}^T + Q_0)^{-1}]^{-1}$$
$$P_{2/1}^{-1} = [\Phi_{2/1}P_1\Phi_{2/1}^T + Q_1]^{-1}$$
$$P_2 = [H_2^T R_2^{-1} H_2 + (\Phi_{2/1}P_1\Phi_{2/1}^T + Q_1)^{-1}]^{-1}$$
$$\cdots\cdots$$

依此类推，显见，随着递推步数的增大，P_0对P_k的影响愈来愈小，即P_k将不依赖于P_0的选取。此性质对于时变或定常模型均适用。

对于完全可观测的和完全可控制的定常信号模型，其卡尔曼滤波器除了具有上述诸稳定性质外，还具有下列性质。

定理6 对于完全可观测的和完全可控制的定常信号模型，若其卡尔曼滤波器是一稳定的线性反馈系统。则当滤波步数充分大之后，滤波误差协方差阵P_k趋于一极限$\overline{P}$，而与初值P_0的选取无关，其中$\overline{P}$为下述矩阵黎卡蒂方程的唯一正定解。

$$\overline{P} = (I - \overline{P}H^T R^{-1} H)(\Phi\overline{P}\Phi^T + Q) \qquad (7\text{-}6\text{-}11)$$

并且增益K_k也趋于极限$\overline{K}$

$$\overline{K}=\overline{P}H^TR^{-1} \tag{7-6-12}$$

证：因为卡尔曼滤波器中P_k的递推关系为

$$P_k=(I-P_kH^TR^{-1}H)(\Phi P_{k-1}\Phi^T+Q) \tag{7-6-13}$$

如果P_k存在极限$\overline{P}$，则必有

$$\lim_{k\to\infty}P_k=\lim_{k\to\infty}P_{k-1}=\overline{P}$$

所以，当$k\to\infty$时，(7-6-13)式就转化成(7-6-11)式。又因为在上一节中已证明P_k存在，且是唯一的正定有界矩阵，所以$\overline{P}$也一定存在，且是(7-6-11)式矩阵黎卡蒂方程的唯一正定解。至于P_k是否存在极限$\overline{P}$，分三种情况证明之。

第一种情况，选取$K_k=\overline{K}$，初值$P_0=\overline{P}$，代入(7-6-13)式可得到

$$P_1=(I-\overline{P}H^TR^{-1}H)(\Phi\overline{P}\Phi^T+Q)=\overline{P}$$

以此类推，有

$$P_0=P_1=\cdots=P_{k-1}=P_k=\overline{P}$$

根据卡尔曼滤波公式

$$P_k=(I-K_kH)(\Phi P_{k-1}\Phi^T+Q)(I-K_kH)^T+K_kRK_k^T$$

$$K_k=P_kH^TR^{-1}$$

把上述结果代入，可知对于任意时刻k，均有

$$\left.\begin{aligned}&\overline{P}=(I-\overline{K}H)(\Phi\overline{P}\Phi^T+Q)(I-\overline{K}H)^T+\overline{K}R\overline{K}^T\\&\overline{K}=\overline{P}H^TR^{-1}\end{aligned}\right\} \tag{7-6-14}$$

第二种情况，先构造一恒增益滤波器，取$K_k=\overline{K}$，$P_0^*>\overline{P}$，则有

$$P_k^*=(I-\overline{K}H)(\Phi P_{k-1}^*\Phi^T+Q)(I-\overline{K}H)^T+\overline{K}R\overline{K}^T \tag{7-6-15}$$

(7-6-15)式减(7-6-14)式，得

$$P_k^*-\overline{P}=(I-\overline{K}H)\Phi(P_{k-1}^*-\overline{P})\Phi^T(I-\overline{K}H)^T \tag{7-6-16}$$

逐步递推，可得

$$P_k^*-\overline{P}=[(I-\overline{K}H)\Phi]^k(P_0^*-\overline{P})[(I-\overline{K}H)\Phi]^{Tk} \tag{7-6-17}$$

由定理2知$\Psi=(I-\overline{K}H)\Phi$的特征值位于单位圆内，故有

$$\lim_{k\to\infty}\Psi^k(P_0^*-\overline{P})(\Psi^T)^k=0$$

所以得

$$\lim_{k\to\infty}P_k^*=\overline{P} \tag{7-6-18}$$

再在卡尔曼滤波器中，仍取P_0^*作为初值，有

$$P_k=(I-K_kH)(\Phi P_{k-1}\Phi^T+Q) \tag{7-6-19}$$

其中K_k为最佳增益，是按照无偏最小方差准则导出的，故必有$P_k<P_k^*$。由定理4得知

$$P_k>\overline{P}，\text{故有}\quad P_k^*>P_k>\overline{P} \tag{7-6-20}$$

既然P_k^*的极限为$\overline{P}$，那么P_k的极限也必然是$\overline{P}$，故有

$$\lim_{k\to\infty}P_k=\overline{P} \tag{7-6-21}$$

第三种情况，若在恒增益滤波器中，选取$K_k=\overline{K}$，$0\leqslant P_0^*<\overline{P}$。此时仍然满足(7-6-17)式，故仍有

$$\lim_{k\to\infty} P_k^* = \overline{P} \tag{7-6-22}$$

但最优滤波器的P_k恒小于次优滤波器的P_k^*，故此时不能看出最优滤波器中的P_k是否收敛于$\overline{P}$。

因为在最优滤波器中，有

$$K_k = P_{k/k-1}H^T(HP_{k/k-1}H^T+R)^{-1}$$
$$= (\Phi P_{k-1}\Phi^T H^T + QH^T)(HQH^T + R + H\Phi P_{k-1}\Phi^T H^T)^{-1} \tag{7-6-23}$$

取$P_0=0$，逐步递推，有

$$HK_1 = HQH^T(HQH^T+R)^{-1}$$
$$HK_2 = (H\Phi P_1\Phi^T H^T + HQH^T)(HQH^T + R + H\Phi P_1\Phi^T H^T)$$
$$HK_3 = (H\Phi P_2\Phi^T H^T + HQH^T)(HQH^T + R + H\Phi P_2\Phi^T H^T)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$

对于一个真分数，若在分子分母上加相同的正数，使分数变大，故K_k是递增的，相应地由$K_k=P_kH^TR^{-1}$易知P_k也是递增的。但P_k有上界P_k^*，而P_k^*的极限为$\overline{P}$，故此P_k的极限也为$\overline{P}$。

对于初值$\overline{P}>P_0>0$的情况，则由定理4易知P_k的极限也为$\overline{P}$。

迄今，我们证明了只要初值P_0取正定阵，不论其值大小，卡尔曼滤波器的P_k均收敛于$\overline{P}$。

下面用一数字例子加以验证。

例7-6-1 设有定常信号模型

$$X_k = 2X_{k-1} + W_{k-1}$$
$$Z_k = X_k + V_k$$

$E[W_k]=0$，$E[W_k^2]=4$，$E[V_k]=0$，$E[V_k^2]=8$

试给出卡尔曼滤波器中P_k的变化规律。

解： 易知此信号模型是完全可观测的和完全可控制的，故此，其卡尔曼滤波器为一稳定的线性系统。在此滤波器中，有

$$P_{k/k-1} = 4P_{k-1} + 4$$

$$K_k = \frac{P_{k-1}+1}{P_{k-1}+3}$$

$$P_k = (1-K_kH)P_{k/k-1} = \frac{8(P_{k-1}+1)}{P_{k-1}+3}$$

求出P_k的稳态解，令$P_{k-1}=P_k=\overline{P}$，得

$$\overline{P} = \frac{8(\overline{P}+1)}{\overline{P}+3}, \quad \overline{P}^2 - 5\overline{P} - 8 = 0$$

解出 $\overline{P} = 6.275,\ -1.275$

取正解 $\overline{P} = 6.275$

从不同初值P_0出发，求P_k，$k=1$，2，…。结果如表7-6-1。

表7-6-1 P_k 的变化规律

P_k / P_0	1	2	3	4	5	∞
∞	8	6.545	6.324	6.284	6.277	6.275
10	6.769	6.382	6.295	6.279	6.276	6.275
6.275	6.275	6.275	6.275	6.275	6.275	6.275
2	4.80	5.940	6.212	6.263	6.273	6.275
0	2.667	5.177	6.043	6.231	6.267	6.275

可见只要递推5步，滤波器就接近稳态。

下面再研究次优滤波器中P_k^*的变化规律。有

$$P_k^* = (1 - \bar{K}H)(\Phi P_{k-1}^* \Phi^T + Q)(1 - \bar{K}H)^T + \bar{K}R\bar{K}^T$$

其中 $\bar{K} = \frac{1}{8} \times 6.275 = 0.784$

故 $P_k^* = 0.216^2(4P_{k-1}^* + 4) + 0.784^2 \times 8$

$= 0.187(P_{k-1}^* + 1) + 4.917$

结果如表7-6-2。

表7-6-2 P_k^* 的变化规律

P_k^* / P_0^*	1	2	3	4	5	∞
10	6.974	6.408	6.302	6.283	6.279	6.275
2	5.478	6.128	6.250	6.272	6.275	6.275
0	5.104	6.058	6.237	6.270	6.275	6.275

定理6的物理解释也是显然的，如果信号模型不是完全可观测的和完全可控制的，那末其卡尔曼滤波器不能保证有界输入Z_k时，输出$\hat{X}_k$有界，也不能保证P_k的正定和有界，此时卡尔曼滤波器是一个不稳定的反馈系统，因而就不可能存在稳态解$\bar{P}$。所以为了保证$\bar{P}$的存在，信号模型应当是定常的，完全可观测的和完全可控制的。

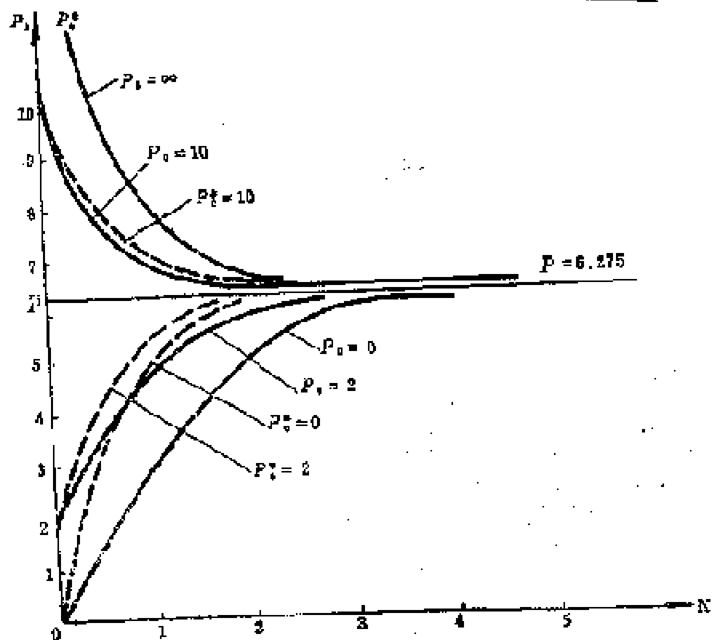


图7-6-1 例7-6-1中P_k、P_k^*的变化规律

另外，由上述例子还可得出一个引理。

引理 对于定常模型，在滤波转移矩阵

$$\Psi_{k/0} = \Psi_{1/0}\Psi_{2/1}\cdots\Psi_{k-1/k-2} \times \Psi_{k/k-1}$$

中，若$\Psi_{i/i-1}$是递降的，即使在$\Psi_{i/i-1}$之前有限项的特征值位于单位圆外，但自$\Psi_{i/i-1}$起其特征值位于单位圆内，则仍能保证卡尔曼滤波器收敛。

证：重写(7-6-17)式，在次优滤波器中有

$$P_k^* - \overline{P} = \Psi^k (P_0^* - \overline{P}) \Psi^{Tk}$$

因为ψ是递降的，当k足够大时，$|L_i(\Psi_{k/k-1})| < 1$，故

$$\lim_{k\to\infty} \Psi_{k/0}^k (P_0^* - \overline{P}) \psi_{k/0}^{Tk} = 0$$

仍然成立，因而有

$$\lim_{k\to\infty} P_k^* = \overline{P}$$

仿照前面的方法，可以证明在最优滤波器中仍有

$$\lim_{k\to\infty} P_k = \overline{P}$$

对于定常信号模型，如果一开始就用稳态增益$\overline{K}$作为滤波器的增益，则滤波器就成了时不变系统，使滤波器结构大大简化。虽然这个滤波器是次优的，但从上述例子可看到，对滤波性能影响不大。所以对于定常信号模型，我们宁愿采用时不变滤波器。

例7-6-2 考察第六章中的α-β-γ滤波器。如果取滤波器的增益常数为稳态值$\overline{\alpha}$、$\overline{\beta}$和$\overline{\gamma}$，那么在时间充分长之后，就可得到状态变量的最佳估计。已知模型为

$$\begin{bmatrix} x \\ \dot{x} \\ \ddot{x} \end{bmatrix}_k = \begin{bmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \\ \ddot{x} \end{bmatrix}_{k-1} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} W_{k-1}$$

$$Z_k = [1,\ 0,\ 0] \begin{bmatrix} x \\ \dot{x} \\ \ddot{x} \end{bmatrix}_k + V_k$$

$$EW_k = 0,\quad E[W_k^2] = Q$$

$$EV_k = 0,\quad E[V_k^2] = R$$

解：先研究一下模型的可观测性与与可控制性。可观测性矩阵为

$$M = \begin{bmatrix} H \\ H\phi \\ H\phi^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & T & \frac{T^2}{2} \\ 1 & 2T & 2T^2 \end{bmatrix}$$

可控制性矩阵为

$$C = [\Gamma,\ \Phi\Gamma,\ \Phi^2\Gamma] = \begin{bmatrix} 0 & \frac{T^2}{2} & 2T^2 \\ 0 & T & 2T \\ 1 & 1 & 1 \end{bmatrix}$$

可见，M和C都满秩，所以此模型是可观测的与可控制的。

由黎卡蒂方程可解出稳态协方差阵$\overline{P}$

$$\overline{P}=\begin{bmatrix}\overline{P}_{11} & \overline{P}_{12} & \overline{P}_{13}\\ \overline{P}_{21} & \overline{P}_{22} & \overline{P}_{23}\\ \overline{P}_{31} & \overline{P}_{32} & \overline{P}_{33}\end{bmatrix}$$

和稳态增益$\overline{K}=\overline{P}H^TR^{-1}$，由于测量值为标量，可推知$\overline{K}$为矢量，记

$$\overline{K}=\begin{bmatrix}\overline{\alpha}\\ \dfrac{\overline{\beta}}{T}\\ \dfrac{2}{T^2}\overline{\gamma}\end{bmatrix}$$

则有
$$\overline{\alpha}=\frac{\overline{P}_{11}}{R},\quad \overline{\beta}=\frac{\overline{P}_{21}}{R}T,\quad \overline{\gamma}=\frac{\overline{P}_{31}}{R}\cdot\frac{T^2}{2}$$

当我们取$\overline{K}$作为α-β-γ滤波器的权系数之后，即使任意选取初值在时间充分长之后，滤波器将成为最佳滤波器。

§7-7 卡尔曼滤波器中的新息与增益

我们重写卡尔曼滤波器方程如下：

$$\hat{X}_{k/k-1}=\Phi_{k/k-1}\hat{X}_{k-1}$$
$$P_{k/k-1}=\Phi_{k/k-1}P_k\Phi^T_{k/k-1}+Q_{k-1}$$
$$\hat{X}_k=\hat{X}_{k/k-1}+K_k(Z_k-H_k\hat{X}_{k/k-1})$$
$$K_k=P_{k/k-1}H_k^T(H_kP_{k/k-1}H_k^T+R_k)^{-1}$$
$$P_k=(1-K_kH_k)P_{k/k-1}$$

考虑滤波方程，记

$$r_k=Z_k-H_k\hat{X}_{k/k-1} \tag{7-7-1}$$

我们称r_k为滤波器的新息，意指在$H_k\hat{X}_{k/k-1}$中包含有状态的老信息，而Z_k中还包含有状态的新信息。

考察一下新息r_k的统计性质。由于Z_k、$\hat{X}_{k/k-1}$都是随机变量，故r_k也是一随机变量。

$$\begin{aligned}r_k&=Z_k-H_k\hat{X}_{k/k-1}\\&=H_kX_k+V_k-H_k\hat{X}_{k/k-1}\\&=H_k\tilde{X}_{k/k-1}+V_k\end{aligned} \tag{7-7-2}$$

其中$\tilde{X}_{k/k-1}=X_k-\hat{X}_{k/k-1}$为预报误差。前面已指出$\hat{X}_{k/k-1}$是$X_k$的一个无偏估计，

故有

$$E[r_k] = H_k E[\widetilde{X}_{k/k-1}] + E[V_k] = 0 \tag{7-7-3}$$

$$E[r_k r_k^T] = E\{(H_k \widetilde{X}_{k/k-1} + V_k)(H_k \widetilde{X}_{k/k-1} + V_k)^T\}$$

$$= H_k P_{k/k-1} H_k^T + R_k \tag{7-7-4}$$

这里应用了V_k与$\widetilde{X}_{k/k-1}$不相关这一结果。

再考察新息的相关矩阵$E[r_k r_j^T]$，$j<k$。因为

$$r_k = Z_k - H_k \Phi_{k/k-1} \hat{X}_{k-1}$$

$$= Z_k - H_k \Phi_{k/k-1}[(I - K_{k-1} H_{k-1})\Phi_{k-1/k-2} \hat{X}_{k-2} + K_{k-1} Z_{k-1}]$$

$$= Z_k - H_k \Phi_{k/k-1} K_{k-1} Z_{k-1} - H_k \Phi_{k/k-1}(I - K_{k-1} H_{k-1})\Phi_{k-1/k-2} \hat{X}_{k-2} \tag{7-7-5}$$

所以，r_k是测量集合$\{Z_k\}$的线性组合。但另一方面，由于$P_{k/k-1}$与测量值集合$\{Z_{k-1}\}$无关系，也就是r_k与集合$\{Z_{k-1}\}$不相关，同理，r_j是集合$\{Z_j\}$的线性组合，且r_j与集合$\{Z_{j-1}\}$不相关。由此可知r_k与r_j必定为不相关，故可以得出结论：当滤波器处于最佳状态时，新息序列为白噪声序列，其均值为零，方差等于$[H_k P_{k/k-1} H_k^T + R_k]$。

新息的这一特征可用来监视滤波器的工作情况。若新息严重偏离白色，表明滤波器工作在非最佳状态。产生不正常的原因有二：一是目标运动产生了强烈机动，原来的状态方程已不适应；二是测量仪器产生故障，引入了大的测量偏差。

检验方法如下：

取5～20个最近的新息样本。令检验统计量为

$$L = \sum_{i=k-N+1}^{k} \frac{r_i^2}{\sigma_i^2} \tag{7-7-6}$$

其中σ_k^2取$[H_k P_{k/k-1} H_k^T + R_k]$的对角线项之和。令$L$与某一门限$G$比较，当$L$持久地超过门限时，表明新息偏量漂移或强度过大，滤波器工作不正常。不正常的新息示例如图7-7-1所示。

再考察滤波器中的增益。因为滤波器工作在最佳状态时，增益为最佳值，最佳增益为

$$K_k = P_{k/k-1} H_k^T [H_k P_{k/k-1} H_k^T + R_k]^{-1}$$

$$= [\Phi_{k/k-1} P_{k-1} \Phi_{k/k-1}^T + Q_{k-1}] H_k^T$$

$$\cdot [H_k(\Phi_{k/k-1} P_{k-1} \Phi_{k/k-1}^T + Q_{k-1}) H_k^T + R_k]^{-1} \tag{7-7-7}$$

假设在$k-1$时刻，滤波器是最佳的，P_{k-1}是最小方差的，由于Q_{k-1}和R_k知道得不准确，代入了错误的值，由此算出的增益K_k就不再是最佳增益了。这时估计误差的协方差阵P_k不应采用公式$P_k = (I - K_k H_k) P_{k/k-1}$，因为这是滤波器最佳时的$P_k$表达式。当增益不为最佳时，应当采用下式求$P_k$。

$$P_k = (I - K_k H_k) P_{k/k-1} (I - K_k H_k)^T + K_k R_k K_k^T \tag{7-7-8}$$

使权矩阵不是最佳的因素有三：1.起始的P_0了解不正确，而用P_0^*代入；2.Q_k了解不正确，而用Q_k^*代入；3.R_k了解不正确，而用R_k^*代入。

上节中已指出，初值P_0不准确的影响会逐渐消失，所以，为了得到最佳滤波效果，应当得到最佳增益，而为了得到最佳增益，应当仔细地选取Q_k和R_k之值，但在许多实际情况下Q_k和R_k是不容易得知的，为了让滤波器尽量接近于最优方式，近来发展了多种自适应卡尔曼滤波算法。

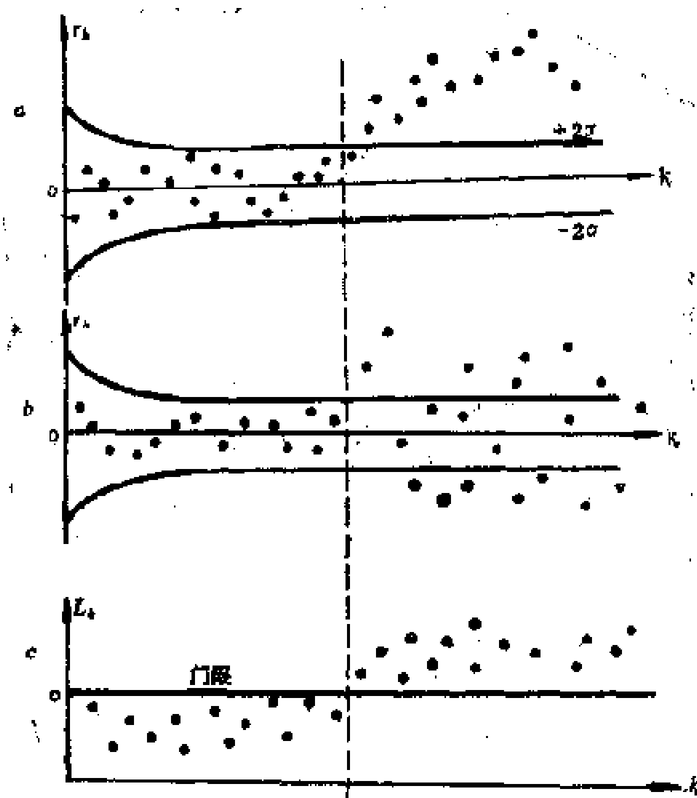


图7-7-1　不正常新息示例

a—新息偏量漂移；　　b—新息强度增大；　　c—新息不正常检验

§7-8　定常信号模型恒增益滤波器的稳定性

对于定常信号模型，其恒增益次优滤波器具有如下的稳定性性质。

定理　对于定常模型的恒增益次优滤波器，若选定恒增益K^*，使状态转移矩阵

$$\Psi^* = (I - K^*H)\Phi \tag{7-8-1}$$

的特征值位于单位圆内，则滤波器是一稳定系统，协方差阵P_k^*收敛于常值阵$\bar{P}^*$，$\bar{P}^*$为下述矩阵方程的正定解。

$$\bar{P}^* = (I - K^*H)(\Phi \bar{P}^*\Phi^T - Q)(I - K^*H)^T + K^*RK^{*T} \tag{7-8-2}$$

若Ψ^*的特征值位于单位圆外或单位圆上，则$\bar{P}^*$不存在，滤波器是一不稳定系统。

证：设有矩阵$\bar{P}^*$满足下述方程

$$\begin{aligned}\bar{P}^* &= (I - K^*H)(\Phi \bar{P}^*\Phi^T + Q)(I - K^*H)^T + K^*RK^{*T} \\ &= \Psi^* \bar{P}^*\Psi^{*T} + (I - K^*H)Q(I - K^*H)^T + K^*RK^{*T}\end{aligned} \tag{7-8-3}$$

对Ψ^*作相似变换，使其对角化，设Λ为$\bar{P}^*$的特征值对角线阵，L为相应的模态矩阵，则有

$$L^{-1}\Psi^*L = \Lambda \tag{7-8-4}$$

$$\begin{aligned}L^{-1}\bar{P}^*L &= L^{-1}\Psi^*\bar{P}^*\Psi^{*T}L + \text{正定阵}A \\ &= L^{-1}\Psi^*LL^{-1}\bar{P}^*LL^{-1}\Psi^{*T}L + \text{正定阵}A \\ &= \Lambda(L^{-1}\bar{P}^*L)\Lambda + \text{正定阵}A\end{aligned} \tag{7-8-5}$$

显然，若$|\lambda_i(\Psi^*)| < 1$，则$\bar{P}^*$为正定阵。

若 $|\lambda_i(\Psi^*)|\geq 1$，则 $\overline{P}^*$ 为负定或不定。因$\overline{P}^*$为协方差阵，必须是正定的，故稳态解 $\overline{P}$ 存在的充分必要条件是要求 $|\lambda_i(\Psi^*)|<1$。

本性质为定常模型恒增益滤波器的增益选择提供了理论依据。

推论 在定常模型恒增益滤波器中，若选取增益$K^*=\overline{K}$，$\overline{K}$为最优卡尔曼滤波器的稳态增益，则恒增益滤波器的稳态方差阵$\overline{P}^*=\overline{P}$，$\overline{P}$为最优卡尔曼滤波器的稳态方差阵。

证：对于最优卡尔曼滤波器而言，最优稳态方差阵$\overline{P}$与$\overline{K}$的关系为

$$\begin{aligned}\overline{P}&=(I-\overline{K}H)(\Phi\overline{P}\Phi^T+Q)\\&=(I-\overline{K}H)(\Phi\overline{P}\Phi^T+Q)(I-\overline{K}H)^T+\overline{K}R\overline{K}^T\end{aligned}\tag{7-8-6}$$

在恒增益滤波器中，当$K^*=\overline{K}$时，其稳态方差阵P^*为

$$\overline{P}^*=(I-\overline{K}H)(\Phi\overline{P}^*\Phi^T+Q)(I-\overline{K}H)^T+\overline{K}R\overline{K}^T\tag{7-8-7}$$

显然，有

$$\overline{P}^*=\overline{P}\tag{7-8-8}$$

所以，若在恒增益滤波器中，选取$K^*=$最优卡尔曼滤波器的稳态增益$\overline{K}$，则此恒增益滤波器具有最小稳态方差。

下面举一数字例子加以验证，仍以例7-6-1标量定常模型为例。

1. 选定$K^*=0.9$，则$\Psi^*=(1-K^*H)\Phi=0.1\times2=0.2<1$，滤波器稳定。由$\overline{P}=0.04\overline{P}^*+6.52$解出稳态值$\overline{P}^*=6.792$，再求不同初值$P_0^*$时的$P_k^*$，$P_k^*$的递推式为$P_k^*=0.04P_{k-1}^*+6.52$。

表7-8-1 P_k^* 值

P_0^*	P_1^*	P_2^*	P_3^*	P_4^*	P_5^*
10	6.920	6.797	6.792	6.792	6.792
6.275	6.771	6.791	6.792	6.792	6.792
2	6.60	6784	6.791	6.792	6.792
0	6.52	6.781	6.791	6.792	6.792

2. 选定$K^*=0.784=\overline{K}$，已知0.784是最优卡尔曼滤波器的稳态增益，由$\Psi^*=(1-K^*H)\Phi=0.216\times2=0.432<1$，得知滤波器稳定，其最优稳态方差为

$$\overline{P}^*=0.784\times8=6.275$$

P_k^*的递推关系式为

$$P_k^*=0.216^2(4P_{k-1}^*+4)+8\times0.784^2=0.186P_{k-1}^*+5.103$$

3. 取$K^*=0.6$，由于$\Psi^*=(1-K^*H)\Phi=0.4\times2=0.8<1$，故滤波器稳定。

由$\overline{P}^*=0.4^2(4\overline{P}^*+4)+0.6^2\times8$解出稳态值$\overline{P}^*=9.778$。

P_k^*的递推关系式为$P_k^*=0.64P_{k-1}^*+3.52$。

4. 取$K^*=0.5$。由$\Psi^*=(1-K^*H)\Phi=0.5\times2=1$，得知$\overline{P}^*$不存在，滤波器为不稳定系统，容易导出$P_k^*$的递推关系式为$P_k^*=P_{k-1}^*+3$，故$P_k^*$是发散的。

5. 取$K^*=0.4$。由$\Psi^*=(1-K^*H)\Phi=0.6\times2=1.2>1$，得知滤波器不稳定。再由$\overline{P}^*=1.44\overline{P}^*+2.72$，得$\overline{P}^*$为负值，故$\overline{P}^*$不存在。

P_k^*的递推式为 $P_k^* = 1.44P_{k-1}^* + 2.72$。

显然，从初值P_0^*出发，P_k^*是发散的，随着k增长而无限增大。

下图中画出了 $K^*=0.9$，$K^*=0.784$, $K^*=0.6$，$K^*=0.5$，$K^*=0.4$ 时和初值$P_0^*=10$，$P_0^*=2$时P_k^*的变化曲线，见图7-8-1。

表7-8-2 P_k^* 值

	P_0^*	P_1^*	P_2^*	P_3^*	P_4^*	P_∞^*
$K^*=0.9$	10	6.920	6.797	6.792	6.792	6.792
	2	6.60	6.784	6.791	6.792	6.792
$K^*=0.784$	10	6.965	6.398	6.295	6.275	6.275
	2	5.475	6.121	6.242	6.265	6.275
$K^*=0.4$	10	17.12	27.37	42.13		∞
	2	5.020	10.21	17.42		∞
$K^*=0.5$	10	13	16	19	22	∞
	2	5	8	11	14	∞
$K^*=0.6$	10	9.92	9.87	9.84	9.815	9.778
	2	5.30	7.33	8.21	8.775	9.778

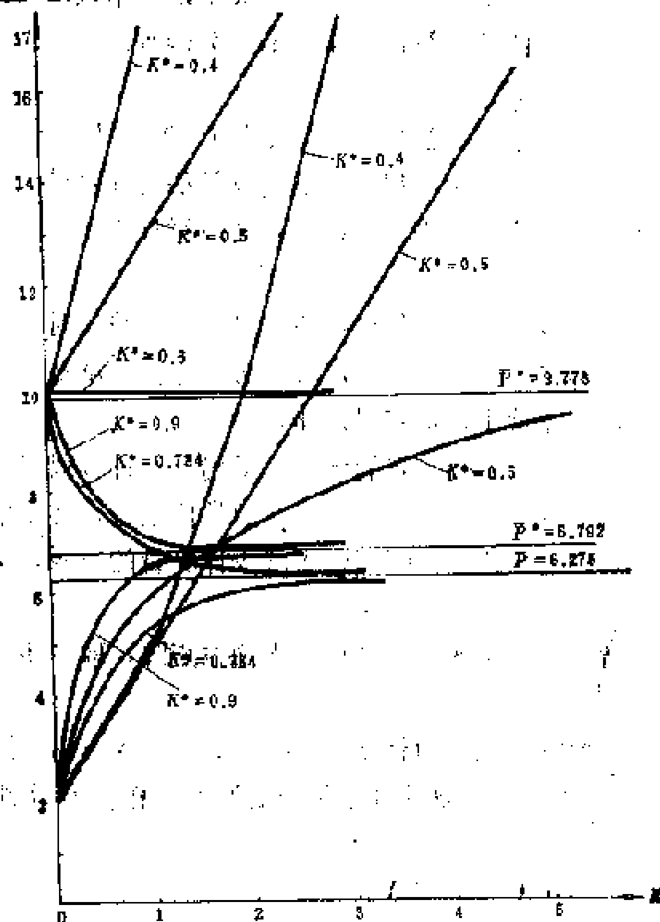


图7-8-1 定常模型恒增益滤波器的P_k^*变化曲线

第八章　卡尔曼滤波的推广

上一章中的卡尔曼滤波是线性无偏最小方差准则下的最优估计，由于它具有递推计算形式、适用于多输入、多输出、非平稳随机过程等许多优点，因而得到广泛应用。但是，为了得到线性无偏最小方差估计，有不少约束条件，如限定状态方程和测量方程是线性的，限定状态噪声和测量噪声是零均值白噪声，它们的方差阵Q_k、R_k为已知等等。

在工程实践中，上述条件往往不满足，为了把卡尔曼滤波的应用加以推广，需要研究不同情况下的卡尔曼滤波的修正问题。本章中将研究下述问题：1．降低滤波计算量的方法、2．有色噪声模型的滤波。3．非线性模型的滤波。4．连续时间模型的离散化。5．连续时间卡尔曼滤波。并简略地介绍卡尔曼平滑。

§8-1　降低滤波计算量的方法

卡尔曼滤波是一种递推算法，每得到一个测量值就进行一次迭代计算，当采样率较高且状态矢量维数较高时，计算量太大，降低采样率又使误差增大。因此，需要研究能够显著降低计算量而精度损失不大的滤波算法。主要有下列两种：

（一）分段循环卡尔曼滤波

工程上经常遇到一种情况是，模型中的各系数矩阵Φ、Γ、H为常值，但噪声强度Q_k、R_k作缓慢变化。对于这种情况，采用常规卡尔曼滤波算法时计算量太大，降低采样率又得不到采样点之间的估计，并且估计误差增大，采用分段循环卡尔曼滤波算法可以降低计算量，同时保证必要的精度。

我们知道，卡尔曼滤波器是一种最优递推滤波器，其增益是时变的，所以这是一种时变滤波器，但滤波器的结构比较复杂，计算量大；另一方面，增益为常值的滤波器是时不变滤波器，其结构简单，计算量小，但不是最佳估计。

在上述两种方法的启发下，吸收二者的优点，得出了一种分段常增益滤波器，或叫分段循环卡尔曼滤波器。它是在一小段时间内取一定的常值增益，各小段的增益不相同，各自根据当时的噪声强度来确定。这样，既照顾了减少计算量的要求，又照顾到噪声强度的变化。

所谓分段循环就是指在一个循环内只计算一次增益矩阵，取N步为一个循环，在这N步内增益矩阵保持不变。

以第一循环为例。在这第一循环内包括$k=0,1,2,\cdots,N-1$，已给出初值$\hat{X}_0$和P_0。

在第一循环的第一步($k=0$)，象常规卡尔曼滤波器一样，计算预报，预报误差方差阵与增益矩阵。

$$\hat{X}_{1/0}=\Phi\hat{X}_0 \quad (8\text{-}1\text{-}1)$$

$$P_{1/0}=\Phi P_0\Phi^T+Q_0 \quad (8\text{-}1\text{-}2)$$

$$K_1 = P_{1/0}H^T[HP_{1/0}H^T + R_0]^{-1} \tag{8-1-3}$$

求得的K_1就作为第一循环中的常增益，在以后各步内(至 $k=N-1$)，只用下列两式递推

$$\hat{X}_{k+1/k} = \Phi \hat{X}_k \tag{8-1-4}$$

$$\hat{X}_{k+1} = \hat{X}_{k+1/k} + K_1(Z_{k+1} - H\hat{X}_{k+1/k}) \tag{8-1-5}$$

一直递推到算出$\hat{X}_N$，第一循环结束，转入第二循环(包括 $k=N$，$N+1$，…，$2N-1$)。

在第二循环的第一步（$k=N$），需要计算新的增益K_{N+1}，以K_{N+1}作为第二循环中的常增益。而要计算K_{N+1}，必先求得上一步的估计误差方差矩阵P_N。现在只有初始方差P_0，要从P_0计算P_N，需要采用一个近似方法。

可以这样设想，即认为从t_0到t_N仅有一步，而在t_N时刻有一等效输入噪声，求t_N时刻的预报量共跨过N步，故有

$$P_{N/0} = \Phi_{N/0}P_0\Phi^T_{N/0} + NQ_0 \tag{8-1-6}$$

这里取NQ_0意指输入噪声按功率相加。

再按照下式计算P_N，

$$P_N = [I - P_{N/0}H^T(HP_{N/0}H^T + \bar{R}_N)^{-1} \cdot H]P_{N/0} \tag{8-1-7}$$

其中 $\bar{R}_N = R_N/N$，取$\bar{R}_N$的理由是考虑到利用了N个测量值，故测量噪声功率减小了N倍。

在取得了P_N之后，就可按下式计算第二循环的增益矩阵K_{N+1}。

$$P_{N+1/N} = \Phi P_N \Phi^T + Q_N \tag{8-1-9}$$

$$K_{N+1} = P_{N+1/N}H^T[HP_{N+1/N}H^T + R_{N+1}]^{-1} \tag{8-1-10}$$

其中Q_N为t_N时刻的输入噪声功率，R_{N+1}为t_{N+1}时刻的测量噪声功率。

在第二循环中，即以K_{N+1}作为常增益进行递推计算。

从 $k=2N$ 起将进入第三个循环，先由P_N求得P_{2N}，再求得K_{2N+1}，以K_{2N+1}作为第三循环的常增益矩阵，以此类推。

这种分段循环卡尔曼滤波算法对于机动缓慢的轨迹估计是比较适用的。分段循环卡尔曼滤波算法的分段方法如图 8-1-1 所示。

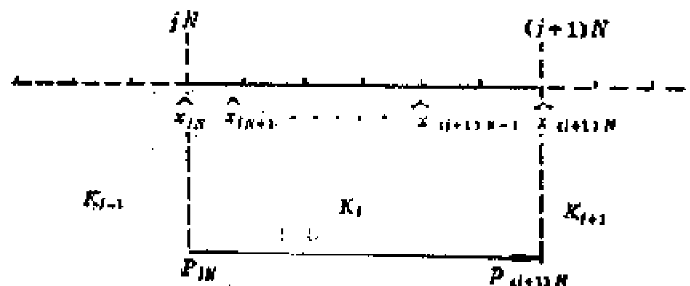


图8-1-1 分段循环卡尔曼滤波器的分段方法

(二) 解耦卡尔曼滤波

解耦卡尔曼滤波，其基本思想是把原高维系统（状态为n维，量测为m维）分解为几个低维的子系统，每个子系统的状态为n_i维，测量为m_i维（$\sum n_i = n$，$\sum m_i = m$），从而分别计算各子系统，可降低计算量若干倍。

首先看在什么条件下，原系统可严格解耦为若干个子系统。目标通常位于三维空间内，可以解耦为三个子系统。

把原状态矢量X与量测矢量Z写成分块形式

$$X=\begin{pmatrix}X^1\\X^2\\X^3\end{pmatrix},\qquad Z=\begin{pmatrix}Z^1\\Z^2\\Z^3\end{pmatrix}\tag{8-1-11}$$

解耦的关键在于计算滤波公式的解耦，现希望把它解耦为

$$\hat{X}^i_{k+1}=\hat{X}^i_{k+1/k}+K^i_{k+1}[Z^i_{k+1}-H^i_{k+1}\hat{X}^i_{k+1/k}]\tag{8-1-12}$$

其中 $H_{k+1}=\begin{pmatrix}H^1_{k+1}\\H^2_{k+1}\\H^3_{k+1}\end{pmatrix}$

如果把增益矩阵K_{k+1}($n\times m$ 维)写成分块形式

$$K_{k+1}=\begin{pmatrix}K^1_{k+1}&K^{12}_{k+1}&K^{13}_{k+1}\\K^{21}_{k+1}&K^2_{k+1}&K^{23}_{k+1}\\K^{31}_{k+1}&K^{32}_{k+1}&K^3\end{pmatrix}\tag{8-1-13}$$

式中K^i_{k+1}对应于第 i 子系统为$n_i\times m_i$维，显然只有在

$$K^{ij}_{k+1}=0,\qquad i\neq j,\qquad i,j=1,2,3\tag{8-1-14}$$

时，上述解耦才成立。也就是说，要求某子系统的滤波值完全不依赖于其它子系统的量测。要使(8-1-14)式成立，又导致了对 $P_{k+1/k}$的要求。

$$P_{k+1/k}=\Phi_{k+1/k}P_k\Phi_{k+1/k}+Q_k\tag{8-1-15}$$

$$K_{k+1}=P_{k+1/k}H^T_{k+1}(H_{k+1}P_{k+1/k}H^T_{k+1}+R_{k+1})^{-1}\tag{8-1-16}$$

选定状态变量与观测量在同一坐标系中。例如三坐标雷达，其测量值 γ、β、ε 通常是独立的，设原状态矢量为$X=[\gamma,\dot{\gamma},\beta,\dot{\beta},\varepsilon,\dot{\varepsilon}]^T$，分解成三个子系统$(\gamma,\dot{\gamma})$、$(\beta,\dot{\beta})$ $(\varepsilon,\dot{\varepsilon})$，此时

$$R_k=\begin{pmatrix}R^1_k&&0\\&R^2_k&\\0&&R^3_k\end{pmatrix}\tag{8-1-17}$$

$$H_k=\begin{pmatrix}1&0&0&0&0&0\\0&0&1&0&0&0\\0&0&0&0&1&0\end{pmatrix}\tag{8-1-18}$$

$$\Phi_{k+1/k}=\begin{pmatrix}\begin{matrix}1&T\\0&1\end{matrix}&&0\\&\begin{matrix}1&T\\0&1\end{matrix}&\\0&&\begin{matrix}1&T\\1&0\end{matrix}\end{pmatrix}\tag{8-1-19}$$

因此R_k、H_k、$\Phi_{k+1/k}$的解耦是容易的，困难的问题是$P_{k+1/k}$的解耦，把 $P_{k+1/k}$写成分块形式

$$P_{k+1/k}=\begin{pmatrix}P^1_{k+1/k}&P^{12}_{k+1/k}&P^{13}_{k+1/k}\\P^{21}_{k+1/k}&P^2_{k+1/k}&P^{23}_{k+1/k}\\P^{31}_{k+1/k}&P^{32}_{k+1/k}&P^3_{k+1/k}\end{pmatrix}\tag{8-1-20}$$

显然，只有在

$$P^{ij}_{k+1/k}=0, \qquad i \neq j \tag{8-1-21}$$

时，系统的解耦才能严格成立。由于目标在运动时，在 r、β、ε 三个方向上互相牵连，因而 $P_{k+1/k}$ 不能严格解耦。但在一定条件下，可近似解耦。下面来讨论近似解耦的条件。

考察 $\hat{X}^1_{k+1}$

$$\begin{aligned}\hat{X}^1_{k+1}=\hat{X}^1_{k+1/k}&+K^1_{k+1}[Z^1_{k+1}-H^1_{k+1}\hat{X}_{k+1/k}]\\&+K^{12}_{k+1}[Z^2_{k+1}-H^2_{k+1}\hat{X}_{k+1/k}]\\&+K^{13}_{k+1}[Z^3_{k+1}-H^3_{k+1}\hat{X}_{k+1/k}]\end{aligned} \tag{8-1-22}$$

若(8-1-12)式中右边最后两项与前两项相比，可忽略不计，则得到解耦后的子系统(8-1-12)式。对于 $\hat{X}^2_{k+1}$、$\hat{X}^3_{k+1}$ 可类推。因此，(8-1-12)式近似成立的条件可表达为均方矩阵的迹满足下述关系：

$$\begin{aligned}&\mathrm{tr}\{K^{is}_{k+1}[H^s_{k+1}P^s_{k+1/k}H^{sT}_{k+1}+R^s_{k+1}]K^{isT}_{k+1}\}\\&<\mathrm{tr}\{\delta^2K^i_{k+1}[H^i_{k+1}P^i_{k+1/k}H^{iT}_{k+1}+R^i_{k+1}]K^{iT}_{k+1}\} \qquad i,s=1,2,3,\ i\neq s\end{aligned} \tag{8-1-23}$$

δ 为正小量，比如可取 $\delta=0.1$。

对于一给定系统，如何判断（8-1-23）式是否成立，可把原系统进行计算机模拟，对结果进行判断。当已作出原系统可近似解耦的判断之后，即可对各子系统分别进行递推计算。

对于 Φ_k 不能写成对角线分块矩阵形式时，即有

$$\Phi_{k+1/k}=\begin{pmatrix}\Phi^{11} & \Phi^{12} & \Phi^{13}\\ \Phi^{21} & \Phi^{22} & \Phi^{23}\\ \Phi^{31} & \Phi^{32} & \Phi^{33}\end{pmatrix}_{k+1/k} \tag{8-1-24}$$

预报估计为

$$\hat{X}^1_{k+1/k}=\Phi^{11}_{k+1/k}\hat{X}^1_k+\Phi^{12}_{k+1/k}\hat{X}^2_k+\Phi^{13}_{k+1/k}\hat{X}^3_k \tag{8-1-25}$$

对于 $\hat{X}^2_{k+1/k}$、$\hat{X}^3_{k+1/k}$ 可类推。

预报误差协方差矩阵为

$$\begin{aligned}P^1_{k+1/k}=\Phi^{11}_{k+1/k}P^1_k\Phi^{11T}_{k+1/k}&+\Phi^{12}_{k+1/k}P^2_k\Phi^{12T}_{k+1/k}\\&+\Phi^{13}_{k+1/k}P^3_k\Phi^{13T}_{k+1/k}+Q^1_k\end{aligned} \tag{8-1-26}$$

这里假设原系统的 Q_k 可近似分解成对角线分块矩阵，Q^1_k 为子系统 1 的状态噪声方差阵。对于 $P^2_{k+1/k}$、$P^3_{k+1/k}$ 可类推。

增益矩阵为

$$K^1_{k+1}=P^1_{k+1/k}H^{1T}_{k+1}[H^1_{k+1}P^1_{k+1/k}H^{1T}_{k+1}+R^1_{k+1}]^{-1} \tag{8-1-27}$$

对于 K^2_{k+1}、K^3_{k+1} 可类推。

滤波估计为

$$\hat{X}^1_{k+1}=\hat{X}^1_{k+1/k}+K^1_{k+1}[Z^1_{k+1}-H_{k+1}\hat{X}^1_{k+1/k}] \tag{8-1-28}$$

对于 $\hat{X}^2_{k+1}$、$\hat{X}^3_{k+1}$ 可类推。

滤波误差协方差阵为

$$P^1_{k+1}=(I-K^1_{k+1}H^1_{k+1})P^1_{k+1/k} \tag{8-1-29}$$

对于 P^2_{k+1}、P^3_{k+1} 可类推。

满足严格解耦的情况通常是很少的，一般地都采用近似解耦，从而滤波器不再是最佳的，多少要损失一些精度，但解耦滤波器的计算量通常要比原系统的滤波器低三倍以上。

§8-2 有色噪声模型的滤波

在卡尔曼滤波器中要求状态噪声和测量噪声都是白噪声，但是许多实际问题不满足这个要求，下面对有色状态噪声情况和有色测量噪声情况分别进行研究。

（一）有色状态噪声时的滤波

常规卡尔曼滤波器中要求状态噪声 w_k 为零均值白噪声序列，但在工程上常常不满足这个要求。例如，一架飞机在作机动飞行时、由于一次机动要在若干个采样周期内完成，因此，其速度扰动不是白噪声序列，而是有色噪声序列。我们希望把卡尔曼滤波器推广应用到有色状态噪声情况，先考察一个例子。

例 8-2-1 设过程的模型为

$$\begin{pmatrix} x_{k+1} \\ \dot{x}_{k+1} \end{pmatrix} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_k \\ \dot{x}_k \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u_k \tag{8-2-1}$$

$$z_{k+1} = [1, \quad 0] \begin{pmatrix} x_{k+1} \\ \dot{x}_{k+1} \end{pmatrix} + v_{k+1} \tag{8-2-2}$$

其中 u_k 是一相关序列，设其相关函数可表达为

$$L_j = E[u_k, u_{j+k}] = \sigma_w^2 \rho^{|j|}, \quad 0 < \rho < 1 \tag{8-2-3}$$

对于有色噪声 u_k 可以看成是白噪声驱动一阶递归滤波器的输出。

解： 对 L_j 作双边 Z 变换，得出 u_j 的功率谱密度为

$$S_u(Z) = \sigma_w^2 \sum_{j=-\infty}^{\infty} \rho^{|j|} Z^{-j}$$

$$= \sigma_w^2 \left\{ \sum_{j=0}^{\infty} \rho^j Z^{-j} + \sum_{j=0}^{\infty} \rho^j Z^j - 1 \right\}$$

$$= \sigma_w^2 \left\{ \frac{1}{1-\rho Z^{-1}} + \frac{1}{1-\rho Z} - 1 \right\}$$

$$= \frac{\sigma_w^2 (1-\rho^2)}{(1-\rho Z)(1-\rho Z^{-1})} \tag{8-2-4}$$

我们把功率谱 $S_u(Z)$ 看成是由白噪声 w_k 激励一线性滤波器所产生，白噪声 w_k 的功率谱密度为 $S_w(Z)$，则有

$$S_w(Z) H(Z) H(Z^{-1}) = S_u(Z) \tag{8-2-5}$$

选取

$$S_w(Z) = \sigma_w^2 (1-\rho^2) \tag{8-2-6}$$

则

$$H(Z) H(Z^{-1}) = \frac{1}{(1-\rho Z)(1-\rho Z^{-1})}$$

$$= \frac{1}{(Z-\rho)(Z^{-1}-\rho)} \tag{8-2-7}$$

滤波器应当是满足因果律的，故选取

$$H(Z)=\frac{1}{Z-\rho} \tag{8-2-8}$$

滤波器 $H(Z)$ 称为成形滤波器。它把白噪声 w_k 变成了有色噪声 u_k。因为

$$W(Z)H(Z)=U(Z)$$

或

$$W(Z)=(Z-\rho)U(Z) \tag{8-2-9}$$

取逆Z变换，得

$$u_{k+1}=\rho u_k+w_k \tag{8-2-10}$$

$\{u_k\}$ 称为一阶马尔科夫序列，产生 $\{u_k\}$ 的滤波器如图8-2-1所示。

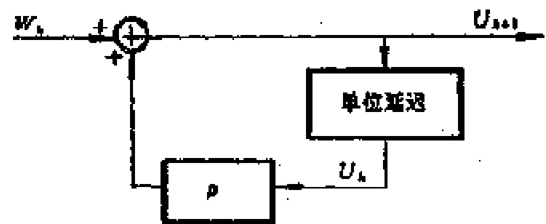


图8-2-1　一阶成形滤波器

因为 u_k 表达了时间过程一个状态的信息，所以我们把 u_k 看成是过程的另一个状态变量，加入到原来的状态矢量中去，于是得出新的状态矢量为

$$X_k=\begin{pmatrix} x_k \\ \dot{x}_k \\ u_k \end{pmatrix} \tag{8-2-11}$$

这时，过程的模型亦应作相应的修改，得

$$\begin{pmatrix} x_{k+1} \\ \dot{x}_{k+1} \\ u_{k+1} \end{pmatrix}=\begin{pmatrix} 1 & T & 0 \\ 0 & 1 & 1 \\ 0 & 0 & \rho \end{pmatrix}\begin{pmatrix} x_k \\ \dot{x}_k \\ u_k \end{pmatrix}+\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}w_k \tag{8-2-12}$$

$$z_{k+1}=[1,0,0]\begin{pmatrix} x_{k+1} \\ \dot{x}_{k+1} \\ u_{k+1} \end{pmatrix}+v_{k+1} \tag{8-2-13}$$

其中 w_k 是白噪声，方差等于 $\sigma_u^2(1-\rho^2)$，于是可以应用常规卡尔曼滤波器求（8-2-12）、(8-2-13)式的最佳估计。

上面介绍的方法称为状态扩充法。下面我们来研究一般形式下的状态扩充卡尔曼滤波器。

设过程的数学模型为

$$X_{k+1}=\Phi_{k+1/k}X_k+W_k \tag{8-2-14}$$

$$Z_{k+1}=H_{k+1}X_k+V_{k+1} \tag{8-2-15}$$

式中 W_k 为马尔科夫序列，可表达为

$$W_k=D_{k/k-1}W_{k-1}+\xi_{k-1} \tag{8-2-16}$$

为白噪声序列，均值为零，方差为 Q_k。应用状态扩充法，令新状态矢量为

$$Y_k=\begin{pmatrix} X_k \\ W_k \end{pmatrix} \tag{8-2-17}$$

则得出新模型为

$$Y_k = \Psi_{k/k-1} Y_{k-1} + \Gamma \xi_{k-1} \tag{8-2-18}$$

$$Z_k = M_k Y_k + V_k \tag{8-2-19}$$

其中

$$\left.\begin{aligned} &\Psi_{k/k-1} = \begin{pmatrix} \Phi_{k/k-1} & I \\ 0 & D_{k/k-1} \end{pmatrix} \\ &\Gamma = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad M_k = [H_k, 0] \end{aligned}\right\} \tag{8-2-20}$$

(8-2-18)、(8-2-19)式满足常规卡尔曼滤波器条件，因此，可以应用卡尔曼滤波公式得到状态扩充卡尔曼滤波器递推公式如下：

预报估计

$$Y_{k/k-1} = \Psi_{k/k-1} \hat{Y}_k \tag{8-2-21}$$

预报误差协方差阵

$$P'_{k/k-1} = \Psi_{k/k-1} P'_{k-1} \Psi^T_{k/k-1} + \Gamma Q_{k-1} \Gamma^T \tag{8-2-22}$$

最佳增益

$$K'_k = P'_{k/k-1} M^T_k [M_k P'_{k/k-1} M^T_k + R_k]^{-1} \tag{8-2-23}$$

滤波估计

$$\hat{Y}_k = \hat{Y}_{k/k-1} + K'_k [Z_k - \hat{M}_k Y_{k/k-1}] \tag{8-2-24}$$

滤波误差协方差阵

$$P'_k = (1 - K'_k M_k) P'_{k/k-1} \tag{8-2-25}$$

然后利用矩阵分块方法，从K'_k、$\hat{Y}_k$、P'_k中分离出所需的K_k、$\hat{X}_k$和P_k。

如果滤波器中的一切参数都能正确给出，那么，由状态扩充法得到的仍然是最优估计。

（二）有色测量噪声时的滤波

许多实际测量噪声并不都是白色的，例如航海多卜勒计程仪测得的航速数据，航空多卜勒雷达测得的飞机航速数据中，都含有有色测量噪声。设信号模型为

$$X_k = \Phi_{k/k-1} X_{k-1} + \Gamma W_{k-1}$$

$$Z_k = H_k X_k + V_k$$

其中V_k为一阶马尔科夫噪声，

$$V_k = \Psi_{k/k-1} V_{k-1} + \eta_{k-1}$$

η_{k-1}为白噪声序列，均值为零，方差为R_{k-1}。

为了应用卡尔曼滤波器，需要把有色噪声转化成白噪声。因为

$$\begin{aligned} Z_{k+1} &= H_{k+1} X_{k+1} + V_{k+1} \\ &= H_{k+1} X_{k+1} + \Psi_{k+1/k} V_k + \eta_k \end{aligned}$$

有

$$Z_{k+1} - \Psi_{k+1/k} Z_k = H_{k+1} X_{k+1} + \eta_k$$

再假定$\Psi_{k+1/k} = \Psi$，则得出新的测量序列

$$(Z_2 - \Psi Z_1), (Z_3 - \Psi Z_2), \cdots, (Z_{i+1} - \Psi Z_i), \cdots$$

当已由初值$\hat{X}_1$和$(Z_2 - \Psi Z_1)$，$(Z_3 - \Psi Z_2), \cdots, (Z_k - \Psi Z_{k-1})$获得$t_k$时刻的估计$\hat{X}_k$，$P_k$之后，滤波器的递推方程为

一步预报

$$\hat{X}_{k+1/k} = \Phi_{k+1/k} \hat{X}_k \tag{8-2-26}$$

预报误差方差阵

$$P_{k+1/k} = \Phi_{k+1/k} P_k \Phi^T_{k+1/k} + \Gamma_k Q_k \Gamma^T_k \tag{8-2-27}$$

滤波估计 $$\hat{X}_{k+1}=\hat{X}_{k+1/k}+K_{k+1}\varepsilon_{k+1} \tag{8-2-28}$$

其中ε_{k+1}表示新息，在本问题中

$$\varepsilon_{k+1}=(Z_{k+1}-H_{k+1}\hat{X}_{k+1/k})-\Psi(Z_k-H_k\hat{X}_k) \tag{8-2-29}$$

K_{k+1}为最优增益，它应当使滤波误差方差矩阵达到最小。

记 $$\tilde{X}_{k+1}=\hat{X}_{k+1}-X_{k+1},\ \tilde{X}_{k+1/k}=\hat{X}_{k+1/k}-X_{k+1} \tag{8-2-30}$$

则有 $$\tilde{X}_{k+1}=\tilde{X}_{k+1/k}+K_{k+1}\varepsilon_{k+1} \tag{8-2-31}$$

取均方，得滤波误差方差矩阵为

$$\begin{aligned}P_{k+1}=&P_{k+1/k}+K_{k+1}E[\varepsilon_{k+1}\varepsilon_{k+1}^T]K_{k+1}^T\\&+E[\tilde{X}_{k+1/k}\varepsilon_{k+1}^T]K_{k+1}^T+K_{k+1}E[\varepsilon_{k+1}\tilde{X}_{k+1/k}^T]\end{aligned} \tag{8-2-32}$$

令 $\dfrac{\partial}{\partial K_{k+1}}(\mathrm{tr}P_{k+1})=0$，求$\mathrm{tr}P_{k+1}$的极小值，得

$$2K_{k+1}E[\varepsilon_{k+1}\varepsilon_{k+1}^T]+2E[\tilde{X}_{k+1/k}\varepsilon_{k+1}^T]=0$$

$$K_{k+1}=-E[\tilde{X}_{k+1/k}\varepsilon_{k+1}^T](E[\varepsilon_{k+1}\varepsilon_{k+1}^T])^{-1} \tag{8-2-33}$$

把ε_{k+1}、$\tilde{X}_{k+1/k}$的表达式代入增益公式

$$\begin{aligned}\varepsilon_{k+1}&=(Z_{k+1}-H_{k+1}\hat{X}_{k+1/k})-\Psi(Z_k-H_k\hat{X}_k)\\&=H_{k+1}X_{k+1}-\Psi H_kX_k+\eta_k-H_{k+1}\hat{X}_{k+1/k}+\Psi H_k\hat{X}_k\\&=-H_{k+1}\tilde{X}_{k+1/k}+\Psi H_k\tilde{X}_k+\eta_k\end{aligned} \tag{8-2-34}$$

$$\tilde{X}_{k+1/k}=\hat{X}_{k+1/k}-X_{k+1}=\Phi_{k+1/k}\tilde{X}_k-\Gamma_kW_k \tag{8-2-35}$$

因此 $$\begin{aligned}\varepsilon_{k+1}&=-H_{k+1}\Phi_{k+1/k}\tilde{X}_k+H_{k+1}\Gamma_kW_k+\Psi H_k\tilde{X}_k+\eta_k\\&=-(H_{k+1}\Phi_{k+1/k}-\Psi H_k)\tilde{X}_k+H_{k+1}\Gamma_kW_k+\eta_k\end{aligned} \tag{8-2-36}$$

已知W_k、η_k为零均值白噪声，且W_k、η_k、$\tilde{X}_k$彼此不相关，故

$$\begin{aligned}E[\varepsilon_{k+1}\varepsilon_{k+1}^T]=&(H_{k+1}\Phi_{k+1/k}-\Psi H_k)P_k(H_{k+1}\Phi_{k+1/k}-\Psi H_k)^T\\&+H_{k+1}\Gamma_kQ_k\Gamma_k^TH_{k+1}^T+R_k\end{aligned} \tag{8-2-37}$$

$$\begin{aligned}E[\tilde{X}_{k+1/k}\varepsilon_{k+1}^T]&=\Phi_{k+1/k}E[\tilde{X}_k\varepsilon_{k+1}^T]-\Gamma_kE[W_k\varepsilon_{k+1}^T]\\&=-\Phi_{k+1/k}P_k(H_{k+1}\Phi_{k+1/k}-\Psi H_k)^T\\&\quad-\Gamma_kQ_k\Gamma_k^TH_{k+1}^T\end{aligned} \tag{8-2-38}$$

于是，测量噪声为一阶自回归模型下的最优增益矩阵K_{k+1}等于

$$\begin{aligned}K_{k+1}=&[\Phi_{k+1/k}P_k(H_{k+1}\Phi_{k+1/k}-\Psi H_k)^T+\Gamma_kQ_k\Gamma_k^TH_{k+1}^T]\\&\times[(H_{k+1}\Phi_{k+1/k}-\Psi H_k)P_k(H_{k+1}\Phi_{k+1/k}-\Psi H_k)^T\\&+H_{k+1}\Gamma_kQ_k\Gamma_k^TH_{k+1}^T+R_k]^{-1}\end{aligned} \tag{8-2-39}$$

在最优增益条件下，ε_{k+1}与$\tilde{X}_{k+1/k}$不相关，故最小滤波误差方差矩阵为

$$\begin{aligned}P_{k+1}&=P_{k+1/k}+K_{k+1}E[\varepsilon_{k+1}\varepsilon_{k+1}^T]K_{k+1}^T\\&=P_{k+1/k}+K_{k+1}E[\varepsilon_{k+1}\tilde{X}_{k+1/k}^T]\end{aligned}$$

$$= P_{k+1/k} - K_{k+1}[(H_{k+1}\Phi_{k+1/k} - \Psi H_k)P_k\Phi_{k+1/k}^T + H_{k+1}\Gamma_k Q_k \Gamma_k^T] \quad (8\text{-}2\text{-}40)$$

滤波估计为

$$\hat{X}_{k+1} = \hat{X}_{k+1/k} + K_{k+1}\varepsilon_{k+1}$$

$$= \hat{X}_{k+1/k} + K_{k+1}[(Z_{k+1} - H_{k+1}\hat{X}_{k+1/k} - \Psi(Z_k - H_k\hat{X}_k)] \quad (8\text{-}2\text{-}41)$$

(8-2-26)、(8-2-27)、(8-2-39)、(8-2-40)、(8-2-41) 诸式构成了最优滤波的递推公式。若令以上诸式中 $\Psi=0$，则还原为一般卡尔曼滤波器。

如果状态噪声与测量噪声均为有色时，则先用状态矢量扩充法重写状态方程，新状态方程与测量方程构成新的信号模型，然后按有色测量噪声情况构造滤波器，

§8-3 非线性模型的滤波

卡尔曼滤波要求信号模型的状态方程和测量方程都是线性的，但是许多工程问题的模型是非线性的。例如，为了较精确地描述目标的运动规律，常常把状态矢量设在直角坐标系中，而测量通常是在极坐标系中进行的，因此信号模型是非线性的。

非线性模型的一般表达式为

$$\left.\begin{aligned} X_{k+1} &= \Phi(X_k) + W_k \\ Z_{k+1} &= h(X_{k+1}) + V_{k+1} \end{aligned}\right\} \quad (8\text{-}3\text{-}1)$$

当然对于一个具体问题，不一定是两个方程都是非线性的。

对于非线性模型的滤波，通常采用广义卡尔曼滤波方法，这是一种次优估计，下面就进行讨论之。

为了利用已有的卡尔曼滤波公式，把非线性函数按泰勒级数展开，取其一阶近似，便可得到线性化的近似信号模型。

例如，把 $\Phi(X_k)$ 在 X_k 邻近按泰勒级数展开，因为 X_k 是不知道的，对 X_k 的最好估计就是 $\hat{X}_k$，所以 $\Phi(X_k)$ 只能围绕 $\hat{X}_k$ 按泰勒级数展开。

$$\Phi(X_k) = \Phi(\hat{X}_k) + \frac{\partial \Phi}{\partial X}\bigg|_{X=\hat{X}_k}(X_k - \hat{X}_k) + \text{高阶项} \quad (8\text{-}3\text{-}2)$$

倘若误差 $\tilde{X}_k = X_k - \hat{X}_k$ 很小，则可以略去高阶项，得到

$$\Phi(X_k) \approx \Phi(\hat{X}_k) + \frac{\partial \Phi}{\partial X}\bigg|_{X=\hat{X}_k}(X_k - \hat{X}_k) \quad (8\text{-}3\text{-}3)$$

定义

$$\Phi_{k+1/k} = \frac{\partial \Phi}{\partial X}\bigg|_{X=\hat{X}_k} \quad (8\text{-}3\text{-}4)$$

$$U_k = \Phi(\hat{X}_k) - \frac{\partial \Phi}{\partial X}\bigg|_{X=\hat{X}_k} \cdot \hat{X}_k \quad (8\text{-}3\text{-}5)$$

代入(8-3-1)式中，得

$$X_{k+1} = \Phi_{k+1/k}X_k + U_k + W_k^* \quad (8\text{-}3\text{-}6)$$

其中 $W_k^* = \Gamma(X_k)\hat{W}_k$ + 略去的高阶项。

注意到W_k^*包含了线性化误差，它依赖于滤波误差$|\hat{X}_k - X_k|$。

同理，把测量方程在X_{k+1}点邻近展开成泰勒级数，因为X_{k+1}是不知道的，这时最好的估计就是$\hat{X}_{k+1/k}$。所以，把$h(X_{k+1})$在$\hat{X}_{k+1/k}$邻近展开成泰勒级数，倘若误差$\tilde{X}_{k+1/k} = X_{k+1} - \hat{X}_{k+1/k}$很小，略去高阶项，则有

$$Z_{k+1} = h(\hat{X}_{k+1/k}) + \frac{\partial h}{\partial X}\bigg|_{X=\hat{X}_{k+1/k}}(X_{k+1} - \hat{X}_{k+1/k}) + V_{k+1} \tag{8-3-7}$$

定义
$$H_{k+1} = \frac{\partial h}{\partial X}\bigg|_{X=\hat{X}_{k+1/k}} \tag{8-3-8}$$

和
$$Y_{k+1} = h(\hat{X}_{k+1/k}) - \frac{\partial h}{\partial X}\bigg|_{X=\hat{X}_{k+1/k}}(\hat{X}_{k+1/k}) \tag{8-3-9}$$

则测量方程成为

$$Z_{k+1} = H_{k+1}X_{k+1} + y_{k+1} + V_{k+1}^* \tag{8-3-10}$$

其中$V_{k+1}^* = V_{k+1}$ + 线性化误差。

此处线性化误差依赖于预报误差$|X_{k+1} - \hat{X}_{k+1/k}|$。

于是，得到线性化后的信号模型为

$$X_{k+1} = \Phi_{k+1/k}X_k + U_k + W_k^*$$
$$Z_{k+1} = H_{k+1}X_{k+1} + y_{k+1} + V_{k+1}^*$$

显然它们完全符合卡尔曼滤波器的要求。我们把U_k看作是已知控制量，把y_{k+1}看作是已知测量偏量，那么，利用前一章的结果，我们可得到相应的卡尔曼滤波器表达式，通常称之为广义卡尔曼滤波器，结果如下：

预报估计

$$\hat{X}_{k+1/k} = \Phi_{k+1/k}\hat{X}_k + U_k = \Phi(\hat{X}_k) \tag{8-3-11}$$

预报误差协方差阵

$$P_{k+1/k} = \Phi_{k+1/k}P_k\Phi_{k+1/k}^T + Q_k^* \tag{8-3-12}$$

增益

$$K_{k+1} = P_{k+1/k}H_{k+1}^T(H_{k+1}P_{k+1/k}H_{k+1}^T + R_{k+1}^*)^{-1} \tag{8-3-13}$$

滤波估计

$$\begin{aligned}\hat{X}_{k+1} &= \hat{X}_{k+1/k} + K_{k+1}[Z_{k+1} - y_{k+1} - H_{k+1}\hat{X}_{k+1/k}] \\ &= \hat{X}_{k+1/k} + K_{k+1}[Z_{k+1} - h(\hat{X}_{k+1/k})]\end{aligned} \tag{8-3-14}$$

滤波误差协方差阵

$$\begin{aligned}P_{k+1} = &(1 - K_{k+1}H_{k+1})P_{k+1/k}(1 - K_{k+1}H_{k+1})^T \\ &+ K_{k+1}R_{k+1}^*K_{k+1}^T\end{aligned} \tag{8-3-15}$$

其中 $Q_k^* = E[W_k^*W_k^{*T}]$

$R_k^* = E[V_k^*V_k^{*T}]$

$$\Phi_{k+1/k} = \frac{\partial \Phi}{\partial X}\bigg|_{X=\hat{X}_k}$$

$$
= \begin{bmatrix} \frac{\partial \Phi_1}{\partial x_1} & \frac{\partial \Phi_1}{\partial x_2} & \cdots & \frac{\partial \Phi_1}{\partial x_n} \\ \frac{\partial \Phi_2}{\partial x_1} & \frac{\partial \Phi_2}{\partial x_2} & \cdots & \frac{\partial \Phi_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial \Phi_n}{\partial x_1} & \frac{\partial \Phi_n}{\partial x_2} & \cdots & \frac{\partial \Phi_n}{\partial x_n} \end{bmatrix}_{X=\hat{X}_k} \tag{8-3-16}
$$

$$
H_{k+1} = \frac{\partial h}{\partial X}\bigg|_{X=\hat{X}_{k+1/k}}
$$

$$
= \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & \cdots & \frac{\partial h_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial h_n}{\partial x_1} & \frac{\partial h_n}{\partial x_2} & \cdots & \frac{\partial h_n}{\partial x_n} \end{bmatrix}_{X=\hat{X}_{k+1/k}} \tag{8-3-17}
$$

对于广义卡尔曼滤波器，应用时应注意以下几点：

1. 在泰勒级数展开式中，只采用一阶近似，因而要求滤波误差 $\tilde{X}_k = X_k - \hat{X}_k$ 和预报误差 $\tilde{X}_{k/k-1} = X_k - \hat{X}_{k/k-1}$ 都很小，保证泰勒级数的一阶近似成立。

2. 由于采用了近似的状态方程和测量方程，不满足线性无偏最小方差条件，所以广义卡尔曼滤波器只是一种次优滤波器，得出的增益不是最佳增益，故采用（8-3-15）式计算 P_k。

3. 广义卡尔曼滤波的效果与初始估计误差 $|X_0 - \hat{X}_0|$ 密切相关，过大的初始误差，将造成滤波的发散。

4. 被舍弃的高阶项可看成是附加噪声，因此在计算增益时，不能用原来的 Q_k、R_k，而应当应用 Q_k^*、R_k^*。通常可取 $Q_k^* = (1+\alpha^k)Q_k$、$R_k^* = (1+\beta^k)R_k$，α、β 为小于 1 的两个常数。

5. 在广义卡尔曼滤波算法中，$P_{k/k-1}$、K_k、P_k 不能脱机计算。

6. 广义卡尔曼滤波器不是最佳滤波器，不能保证滤波的稳定，因此在运行中应当随时加以检验，一种办法是检验新息偏离白色的程度，另一种办法是检验 $\mathrm{tr}P_k$，当检验量不超过某个门限时，就认为滤波器工作正常。

例8-3-1 设有信号模型为

$$x_k = a x_{k-1} + w_{k-1}$$

$$z_k = x_k + v_k$$

这是一种比较常见的情况，信号是一阶马尔科夫序列，但相关系数 a 为未知，假定 w_k、v_k 为互不相关的零均值白噪声序列，其方差分别是 Q_k 与 R_k。求信号 x_k 的估计。

解： 我们把参数 a 看成是信号的一个扩充状态，因为 a 是常数，满足方程

$$a_k = a_{k-1}$$

于是构造状态矢量和状态方程如下：

$$\begin{pmatrix} x_k \\ a_k \end{pmatrix} = \begin{pmatrix} a_{k-1}x_{k-1} \\ a_{k-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} w_k$$

写成矢量形式，得

$$X_k = \Phi(X_{k-1}) + \Gamma w_{k-1}$$

测量方程为

$$z_k = [1, 0]\begin{pmatrix} x_k \\ a_k \end{pmatrix} + v_k$$

$$= HX_k + v_k$$

可见扩充的状态方程是非线性的，我们用广义卡尔曼滤波方法可以同时求得 x_k 与 a_k 的估计。可以求得

$$\frac{\partial \Phi(X_k)}{\partial X_k} = \begin{pmatrix} a_k & x_k \\ 0 & 1 \end{pmatrix}$$

利用(8-3-11)至(8-3-15)式，可求得全部结果如下：

$$\hat{x}(k) = \hat{a}(k-1)\hat{x}(k-1) + K_x(k)[z(k) - \hat{a}(k-1)\hat{x}(k-1)]$$

$$\hat{a}(k) = \hat{a}(k-1) + K_a(k)[z(k) - \hat{a}(k-1)\hat{x}(k-1)]$$

$$K_x(k) = \frac{P_x(k/k-1)}{P_x(k/k-1) + R(k)}$$

$$K_a(k) = \frac{P_{ax}(k/k-1)}{P_x(k/k-1) + R(k)}$$

$$P_x(k/k-1) = \hat{a}^2(k-1)P_x(k-1) + 2\hat{a}(k-1)x(k-1)P_{ax}(k-1)$$

$$+ \hat{x}^2(k-1)P_a(k-1) + Q(k)$$

$$P_{ax}(k/k-1) = \hat{a}(k-1)P_{ax}(k-1) + \hat{x}(k-1)P_a(k-1)$$

$$P_a(k/k-1) = P_a(k-1)$$

$$P_x(k) = \frac{P_x(k/k-1)R(k)}{P_x(k/k-1) + R(k)}$$

$$P_{ax}(k) = \frac{P_{ax}(k/k-1)R(k)}{P_{ax}(k/k-1) + R(k)}$$

$$P_a(k) = P_a(k/k-1) - \frac{P_{ax}^2(k/k-1)}{P_x(k/k-1) + R(k)}$$

标称轨线法

设有非线性模型

$$X_{k+1} = \Phi(X_k) + \Gamma(X_k)W_k \qquad (8\text{-}3\text{-}18)$$

$$Z_{k+1} = h(X_{k+1}) + V_{k+1} \qquad (8\text{-}3\text{-}19)$$

式中 Φ、Γ、h 均是 X_k的非线性函数。假定已具有标称轨线的先验知识，所谓标称轨线是指没有噪声 W_k、V_k时的理想轨线，它们表示为

$$\overline{X}_{k+1} = \Phi(\overline{X}_k) \qquad (8\text{-}3\text{-}20)$$

$$\overline{Z}_{k+1} = h(\overline{X}_{k+1}) \qquad (8\text{-}3\text{-}21)$$

标称轨线一般是在理想条件下通过物理定理推算出来的。由于各种随机干扰的影响，使得

实际轨线围绕标称轨线起伏，我们的目的是要对实际轨线作出好的估计。

令系统状态与标称轨线的偏离为

$$\left.\begin{aligned}\Delta X_k &= X_k-\overline{X}_k\\ \Delta Z_k &= Z_k-\overline{Z}_k\end{aligned}\right\}\tag{8-3-22}$$

当ΔX_k、ΔZ_k足够小时，在标称轨线附近可把$\Phi(X_k)$、$h(X_k)$展开成泰勒级数，略去高次项，以及进一步假定$\Gamma(X_k)=\Gamma(\overline{X}_k)$，则信号模型可以近似地写成

$$X_{k+1}=\Phi(\overline{X}_k)+\frac{\Phi\partial}{\partial X}\bigg|_{X=\overline{X}_k}\cdot\Delta X_k+\Gamma(\overline{X}_k)W_k\tag{8-3-23}$$

$$Z_k=h(X_k)+\frac{\partial h}{\partial X}\bigg|_{X=\overline{X}_k}\cdot\Delta X_k+V_k\tag{8-3-24}$$

于是得到偏离量方程

$$\Delta X_{k+1}=\frac{\partial\Phi}{\partial X}\bigg|_{X=\overline{X}_k}\cdot\Delta X_k+\Gamma(\overline{X}_k)W_k\tag{8-3-25}$$

$$\Delta Z_k=\frac{\partial h}{\partial X}\bigg|_{X=\overline{X}_k}\cdot\Delta X_k+V_k\tag{8-3-26}$$

令

$$\Phi_{\Delta(k+1/k)}=\frac{\partial\Phi}{\partial X}\bigg|_{X=\overline{X}_k}$$

$$H_{\Delta(k+1/k)}=\frac{\partial h}{\partial X}\bigg|_{X=\overline{X}_k}$$

则可写出偏离量的信号模型

$$\Delta X_{k+1}=\Phi_{\Delta(k+1/k)}X_k+\Gamma(\overline{X}_k)\tag{8-3-27}$$

$$\Delta Z_k=H_{\Delta(k)}\Delta X_k+V_k\tag{8-3-28}$$

显然(8-3-27)、(8-3-28)为线性模型，故以ΔX_k为状态变量，以ΔZ_k为测量序列，利用已有的线性卡尔曼滤波器得到递推算法如下：

$$\begin{aligned}
\Delta\hat{X}_{k+1/k}&=\Phi_{\Delta(k+1/k)}\hat{X}_k\\
P_{\Delta(k+1/k)}&=\Phi_{\Delta(k+1/k)}P_{\Delta(k)}\Phi^T_{\Delta(k+1/k)}\\
&\quad+\Gamma(\overline{X}_k)Q_k\Gamma^T(\overline{X}_k)\\
K_{\Delta(k+1)}&=P_{\Delta(k+1/k)}H^T_{\Delta(k+1)}[H_{\Delta(k+1)}P_{\Delta(k+1/k)}H^T_{\Delta(k+1)}+R_{k+1}]^{-1}\\
\Delta\hat{X}_{k+1}&=\Delta\hat{X}_{k+1/k}+K_{\Delta(k+1)}[\Delta Z_{k+1}-H_{\Delta(k+1)}\Delta\hat{X}_{k+1/k}]\\
P_{\Delta(k+1)}&=[I-K_{\Delta(k)}H_{\Delta(k+1)}]P_{\Delta(k+1/k)}
\end{aligned}$$

初始条件　　$\Delta\hat{X}_0=0$，　$P_{\Delta(0)}=P_0$

于是得到实际轨线估计为

$$\hat{X}_k=\overline{X}_k+\Delta\hat{X}_k$$

$$P_k=P_{\Delta(k)}$$

即实际轨线估计是由标称轨线加上修正量组成，标称轨线是理想的，故估计轨线的方差就等于偏离量的方差。应当指出，上述算法只是在随机扰动比较小时才有效，因此，求标称轨线时应尽可能精确些。

§8-4 卡尔曼平滑简述

平滑估计的含义是利用k时刻之前和k时刻之后的观测数据来获得k时刻的状态估计。平滑估计比滤波估计利用了更多的观测值，预料平滑估计误差方差将小于滤波估计误差方差。卡尔曼平滑一般用于事后处理。

平滑有三种方式：

1. 固定区间平滑。对于固定区间$t_1, \cdots, t_k, \cdots, t_j$，根据全部观测值$z_1, \cdots, z_k, \cdots, z_j$，作出每一时刻的状态估计$\hat{X}_{1/j}, \cdots, \hat{X}_{k/j}, \cdots, \hat{X}_{j/j}$。固定区间平滑应用于事后处理，希望获得比卡尔曼滤波更好的估计。

2. 固定点平滑。是指被估计的时刻k是固定的。当我们获得了X_k的平滑值$\hat{X}_{k/j}$之后，再由新测量值Z_{j+1}去求得新的平滑值$\hat{X}_{k/j+1}$，此时递推的初值是$\hat{X}_k$，其递推顺序是

$$\hat{X}_k + Z_{k+1} \longrightarrow \hat{X}_{k/k+1} + Z_{k+2} \longrightarrow \cdots \hat{X}_{k/j} + Z_{j+1} \longrightarrow \hat{X}_{k/j+1}$$

固定点平滑应用于下述情况，比如，当对导弹的自由飞行段进行了跟踪观测和滤波，为了考察制导的精度，要求得到从机动段到自由飞行段转换时刻的导弹位置和速度的精确估计，此时就要采用固定点平滑。

3. 固定滞后时间平滑。即要求各时刻状态的平滑值$\hat{X}_{k/j}$，其滞后时间$L = j - k$是固定的，这里的问题是当获得了平滑值$\hat{X}_{k/j}(j - k = L)$之后，如何由新测量值Z_{k+1}去求得新平滑值$\hat{X}_{k+1/j+1}$，此时的递推初值是由固定点平滑所获得的初始状态的平滑值$\hat{X}_{0/L}$，其递推顺序为

$$\hat{X}_{0/L} + Z_1 \longrightarrow \hat{X}_{1/L+1} + Z_2 \longrightarrow \cdots \longrightarrow \hat{X}_{k/j} + Z_{k+1} \longrightarrow \hat{X}_{k+1/j+1}$$

$$L = j - k$$

如果在上述导弹制导问题中，不仅要得到转换时刻导弹位置、速度的精确估计，还要求得到自由飞行段的位置和速度的精确估计，此时应采用固定时间滞后平滑。

由于后两种平滑方式只在少数情况下得到应用，这里不作讨论。下面仅对固定区间平滑进行考察。

今考察一线性离散模型

$$X_{k+1} = \Phi_{k+1/k} X_k + \Gamma_k W_k$$
$$Z_{k+1} = H_{k+1} X_{k+1} + V_{k+1}$$

其中W_k、V_k为零均值白噪声，且互相独立。

我们试图在t_j时刻（$j > k$）作出估计$\hat{X}_k$，求$\hat{X}_k$可以分三步进行，并用符号$Z(1, k)$表示t_k之前的测量值$\{Z_1, Z_2, \cdots, Z_k\}$，用符号$Z(k+1, j)$表示$t_k$之后的测量值$\{Z_{k+1}, Z_{k+2}, \cdots, Z_j\}$。

第一步，利用测量值$Z(1, k)$，按常规卡尔曼滤波器去求出$\hat{X}_k$，用初值$\hat{X}_0$、P_0启动滤波器，这个滤波器称为前向滤波器。

第二步，利用t_j至t_{k+1}的测量值计算出t_{k+1}时刻的估计$\hat{X}_{k+1}$，再作一步预报至t_k时刻，它的递推顺序是从后往前，称之为后向滤波器，得到的估计用下述符号表示

$$\hat{X}_{bk/k+1} = E[X_k / Z(j, k+1)]$$

由于我们没有关于X_j的先验统计知识，所以后向滤波器以采用逆协方差形式的卡尔

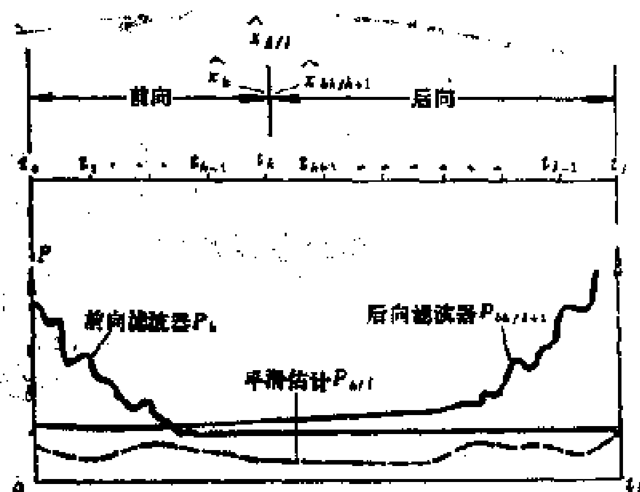


图8-4-1 求平滑估计$\hat{X}_{k/l}$

曼滤波器比较适宜。

第三步，把前向滤波器得到的$\hat{X}_k$和后向滤波器得到的$\hat{X}_{bk/k+1}$，以最佳方式加以组合，根据加权最小二乘估计的定义，得知独立无偏测量的最佳组合就是加权最小二乘估计。所以，$\hat{X}_k$和$\hat{X}_{bk/k+1}$的加权最小二乘估计就是最佳平滑估计。注意不要把Z_k这个数据同时加入到两个滤波器中。

求平滑估计时数据的利用方式及估计方差的改善情况示于图8-4-1中。

§8-5 连续时间模型的离散化

一个动态系统的运动规律，通常可用微分方程加以描述，而测量一般是在离散时间上进行的，即信号模型的状态方程是连续时间形式的，而测量方程是离散时间形式的，为了应用前面的卡尔曼滤波器，为了便于计算机实施，需要把状态方程离散化。

（一）连续定常模型状态方程的离散化

设状态方程为

$$\dot{X}(t)=AX(t)+BW(t) \tag{8-5-1}$$

上述矢量微分方程的解为

$$X(t)=e^{A(t-t_0)}X(t_0)+\int_{t_0}^{t}e^{A(t-\tau)}BW(\tau)d\tau \tag{8-5-2}$$

在上式中取$t_0=kT$，$t=(k+1)T$，得

$$X(kT+T)=e^{AT}X(kT)+\int_{kT}^{(k+1)T}e^{A(kT+T-\tau)}BW(\tau)d\tau$$

$$=e^{AT}X(kT)+\int_0^{T}e^{A(T-\zeta)}BW(kT+\zeta)d\zeta \tag{8-5-3}$$

若输入函数$W(t)$是梯阶形函数，即在每个采样区间内存在关系式

$$W(kT+\zeta)=W(kT)\qquad k=0,1,2,\cdots \tag{8-5-4}$$

则有

$$X(kT+T)=e^{AT}X(kT)+\left[\int_0^{T}e^{A(T-\zeta)}Bd\zeta\right]W(kT) \tag{8-5-5}$$

记

$$\Phi_{k+1/k}=e^{AT}=\Phi \tag{8-5-6}$$

$$\Gamma=\int_0^{T}e^{A(T-\zeta)}Bd\zeta=\int_0^{T}e^{At}Bdt \tag{8-5-7}$$

则得离散化状态方程为

$$X(kT+T)=\Phi X(kT)+\Gamma W(kT) \tag{8-5-8}$$

或

$$X_{k+1}=\Phi X_k+\Gamma W_k \tag{8-5-9}$$

倘若输入$W(t)$不是阶梯形函数，那么，(8-5-8)式只是(8-5-1)式的近似。

由矩阵指数的定义，得

$$e^{AT}=I+AT+\frac{1}{2!}(AT)^2+\frac{1}{3!}(AT^3)+\cdots \tag{8-5-10}$$

直接计算 e^{AT} 是很麻烦的，还难以得到封闭形式的结果。下面介绍另一种计算 e^{AT} 的有效方法。把 e^{AT} 写成下列形式

$$e^{AT}=a_0(T)I+a_1(T)A+\cdots+a_{n-1}(T)A^{n-1} \tag{8-5-11}$$

其中 n 表示矩阵 A 的阶数，如果 A 的各特征值互不相同，有特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$，把各特征值代入下面方程中

$$e^{\lambda_i T}=a_0(T)+a_1(T)\lambda_i+\cdots+a_{n-1}(T)\lambda_i^{n-1} \tag{8-5-12}$$

得到 n 个联立方程，即可算出参数 $a_0(T)$，$a_1(T)$，$\cdots$，$a_{n-1}(T)$。

例8-5-1　设 $A=\begin{pmatrix}0 & 1\\ 0 & 2\end{pmatrix}$，求 e^{AT}。

解：A 的特征方程为

$$|\lambda I-A|=\begin{vmatrix}\lambda & -1\\ 0 & \lambda+2\end{vmatrix}=\lambda(\lambda+2)=0$$

解出　$\lambda_1=0$，$\lambda_2=-2$

把 λ_1、λ_2 分别代入方程

$$e^{\lambda_i T}=a_0(T)+a_1(T)\lambda_i \qquad i=1,2$$

可得

$$1=a_0(T)$$

$$e^{-2T}=a_0(T)-2a_1(T)$$

解出

$$a_0(T)=1, \qquad a_1(T)=\frac{1}{2}(1-e^{-2T})$$

故有

$$e^{AT}=I+\frac{1}{2}(1-e^{-2T})A$$

$$=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}+\frac{1}{2}(1-e^{-2T})\begin{pmatrix}0 & 1\\ 0 & -2\end{pmatrix}$$

$$=\begin{pmatrix}1 & \frac{1}{2}(1-e^{-2T})\\ 0 & e^{-2T}\end{pmatrix}$$

例8-5-2　设有连续时间状态方程

$$\begin{pmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{pmatrix}=\begin{pmatrix}0 & 1\\ 0 & -2\end{pmatrix}\begin{pmatrix}x_1(t)\\ x_2(t)\end{pmatrix}+\begin{pmatrix}0\\ 1\end{pmatrix}w(t)$$

求上式的离散化状态方程。

解：由(8-5-6)式得

$$\Phi=e^{AT}=\begin{pmatrix}1 & \frac{1}{2}(1-e^{-2T})\\ 0 & e^{-2T}\end{pmatrix}$$

由(8-5-7)式得

$$
\Gamma=\left[\int_0^T e^{At}dt\right]B
$$

$$
=\left(\int_0^T\begin{bmatrix}1 & \frac{1}{2}(1-e^{-2t})\\ 0 & e^{-2t}\end{bmatrix}dt\right)\begin{bmatrix}0\\1\end{bmatrix}
$$

$$
=\begin{bmatrix}\int_0^T dt & \int_0^T\frac{1}{2}(1-e^{-2t})dt\\ 0 & \int_0^T e^{-2t}dt\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}
$$

$$
=\begin{bmatrix}T & \frac{1}{2}\left[T+\frac{1}{2}(e^{-2T}-1)\right]\\ 0 & \frac{1}{2}(1-e^{-2T})\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}
$$

$$
=\begin{bmatrix}\frac{T}{2}+\frac{1}{4}(e^{-2T}-1)\\ \frac{1}{2}(1-e^{-2T})\end{bmatrix}
$$

于是，得离散时间状态方程为

$$
\begin{bmatrix}x_1(kT+T)\\ x_2(kT+T)\end{bmatrix}=\begin{bmatrix}1 & \frac{1}{2}(1-e^{-2T})\\ 0 & e^{-2T}\end{bmatrix}\begin{bmatrix}x_1(kT)\\ x_2(kT)\end{bmatrix}+\begin{bmatrix}\frac{T}{2}+\frac{1}{4}(e^{-2T}-1)\\ \frac{1}{2}(1-e^{-2T})\end{bmatrix}w(kT)
$$

当矩阵A的特征值有重根的情况，这时方程组（8-5-12）的个数不够n，不能把系数$a_0(T)$，$a_1(T)$，…，$a_{n-1}(T)$确定出来，这时需要用下面的方法。

若λ_1为三重根，代入(8-5-12)式，得

$$
e^{\lambda_1 T}=a_0(T)+a_1(T)\lambda_1+\cdots+a_{n-1}(T)\lambda_1^{n-1} \tag{8-5-13}
$$

对λ_1求导一次，得

$$
Te^{\lambda_1 T}=a_1(T)+2a_2(T)\lambda_1+\cdots+(n-1)a_{n-1}(T)\lambda_1^{n-2} \tag{8-5-14}
$$

再对λ_1求导一次，得

$$
T^2e^{\lambda_1 T}=2a_2(T)+6a_3(T)\lambda_1+\cdots+(n-1)(n-2)a_{n-1}(T)\lambda_1^{n-3} \tag{8-5-15}
$$

因而仍然得到n个独立方程，就可解出系数$a_0(T)$，$a_1(T)$，…，$a_{n-1}(T)$。

例8-5-3 已知

$$
A=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 2 & 3 & 0\end{bmatrix},
$$

求e^{AT}。

解： 由于A的特征值为

$$
\lambda_1=2,\quad \lambda_2=\lambda_3=-1
$$

得到方程组为

$$
e^{2T}=a_0(T)+2a_1(T)+4a_2(T)
$$

$$e^{-T} = a_0(T) - a_1(T) + a_2(T)$$

和
$$Te^{-T} = a_1(T) - 2a_2(T)$$

解联立方程组，得

$$a_0(T) = \frac{1}{9}(e^{2T} + 8e^{-T} + 6T^2 e^{-T})$$

$$a_1(T) = \frac{1}{9}(2e^{2T} - e^{-T} + 3Te^{-T})$$

$$a_2(T) = \frac{1}{9}(e^{2T} - e^{-T} - 3Te^{-T})$$

于是，得$e^{AT} = a_0(T)\mathbf{I} + a_1(T)\boldsymbol{A} + a_2(T)\boldsymbol{A}^2 = \Phi$

$$= \begin{pmatrix} e^{2T} + (8+6T)e^{-T} & 2e^{2T} - (2-3T)e^{-T} & e^{2T} - (1-3T)e^{-T} \\ 2e^{2T} - (2+6T)e^{-T} & 4e^{2T} + (5-3T)e^{-T} & 2e^{2T} - (2-3T)e^{-T} \\ 4e^{2T} + (6T-4)e^{-T} & 8e^{2T} + (3T-8)e^{-T} & 4e^{2T} + (5-3T)e^{-T} \end{pmatrix}$$

（二）线性非定常状态方程的离散化

设状态方程为

$$\dot{\boldsymbol{X}}(t) = \boldsymbol{A}(t)\boldsymbol{X}(t) + \boldsymbol{B}(t)\boldsymbol{W}(t) \tag{8-5-16}$$

对于上述状态方程的离散化，通常采用近似方法，利用近似关系

$$\dot{\boldsymbol{X}}(kT) \approx \frac{1}{T}[\boldsymbol{X}(kT+T) - \boldsymbol{X}(kT)] \tag{8-5-17}$$

也就是在区间$(k+1)T \sim kT$内，把$\dot{\boldsymbol{X}}(t)$看成是常数，代入(8-5-9)式中，得

$$\frac{1}{T}[\boldsymbol{X}(kT+T) - \boldsymbol{X}(kT)] = \boldsymbol{A}(kT)\boldsymbol{X}(kT) + \boldsymbol{B}(kT)\boldsymbol{W}(kT)$$

或
$$\boldsymbol{X}(kT+T) = [\mathbf{I} + T\boldsymbol{A}(kT)]\boldsymbol{X}(kT) + T\boldsymbol{B}(kT)\boldsymbol{W}(kT)$$

记
$$\Phi_{k+1/k} = \mathbf{I} + T\boldsymbol{A}(kT) \tag{8-5-18}$$

$$\Gamma_k = T\boldsymbol{B}(kT) \tag{8-5-19}$$

则有离散化状态方程为

$$\boldsymbol{X}_{k+1} = \Phi_{k+1/k}\boldsymbol{X}_k + \Gamma_k \boldsymbol{W}_k \tag{8-5-20}$$

只有在T较小时，才有较好的近似度，求严格的离散化表达式相当麻烦，这里从略。

（三）连续时间非线性模型的离散化

在许多情况下，状态方程是连续时间非线性型的。状态方程如下：

$$\frac{d}{dt}\boldsymbol{X}(t) = f[\boldsymbol{X}(t)] + \Gamma(t)\boldsymbol{W}(t) \tag{8-5-21}$$

其中$\boldsymbol{W}(t)$为模型噪声矢量，假定为零均值白噪声，其方差矩阵为

$$E[\boldsymbol{W}(t)\boldsymbol{W}(\tau)^T] = Q(t)\delta(t-\tau) \tag{8-5-22}$$

测量方程是时间离散非线性型的，测量方程为

$$\boldsymbol{Z}_{k+1} = h(\boldsymbol{X}_{k+1}) + \boldsymbol{V}_{k+1} \tag{8-5-23}$$

$\boldsymbol{V}_{k+1}$为零均值白噪声，其方差阵为

$$E[\boldsymbol{V}_{k+1}\boldsymbol{V}_{k+1}^T] = \boldsymbol{R}_{k+1} \tag{8-5-24}$$

现在把(8-5-21)式离散化，对(8-5-21)式取积分，得

$$X_{k+1}-X_k=\int_{t_k}^{t_{k+1}} f[X(t)]dt+\int_{t_k}^{t_{k+1}}\Gamma(t)W(t)dt \tag{8-5-25}$$

当采样间隔T较小时，在区间(t_k, t_{k+1})内可把$f[X(t)]$展开，取其一阶项

$$f[X(t)]=f(X_k)+\frac{\partial f}{\partial X}\frac{\partial X}{\partial t}\bigg|_{X=X_k}(t-t_k) \tag{8-5-26}$$

其中　$X_k=X(t_k)$

$$A(X_k)=\frac{\partial f(X)}{\partial X}\bigg|_{X=X_k} \tag{8-5-27}$$

代入(8-5-26)式中，并在t_k至t_{k+1}区间积分，得

$$X_{k+1}-X_k=f(X_k)T+A(X_k)f(X_k)\frac{T^2}{2}+\int_{t_k}^{t_{k+1}}\Gamma(t)W(t)dt \tag{8-5-28}$$

上式右边第三项为模型噪声，其方差阵为

$$E\left[\int_{t_k}^{t_{k+1}}\Gamma(t)W(t)dt\right]\left[\int_{t_k}^{t_{k+1}}\Gamma(t')W(t')dt'\right]^T$$

$$=\int_{t_k}^{t_{k+1}}\int_{t_k}^{t_{k+1}}\Gamma(t)Q(t)\delta(t-t')\Gamma(t')^T dt'dt \tag{8-5-29}$$

注意到　$\int_{t_k}^{t_{k+1}}\int_{t_k}^{t_{k+1}}Q(t)\delta(t-t')dt\,dt'=\int_{t_k}^{t_{k+1}}Q(t')dt'=Q(t_k)T$

记　　　$\Gamma_k=\Gamma(t_k)$，　　$Q_k=Q(t_k)T$

则上式近似等于$\Gamma_k Q_k \Gamma_k^T$

记　　　$E[(\Gamma_k W_k)(\Gamma_k W_k)^T]=\Gamma_k Q_k \Gamma_k^T$　　(8-5-30)

得出离散化状态方程为

$$X_{k+1}=X_k+f(X_k)T+A(X_k)f(X_k)\frac{T^2}{2}+\Gamma_k W_k \tag{8-5-31}$$

若已知k时刻估计$\hat{X}_k$，取一步预报

$$\hat{X}_{k+1/k}=\hat{X}_k+f(\hat{X}_k)T+A(\hat{X}_k)f(\hat{X}_k)\frac{T^2}{2}\approx\hat{X}_k+f(\hat{X}_k)T \tag{8-5-32}$$

预报误差为

$$\begin{aligned}\tilde{X}_{k+1/k}&=X_{k+1}-\hat{X}_{k+1/k}\\&=X_k-\hat{X}_k+[f(X_k)-f(\hat{X}_k)]T+\Gamma_k W_k+\xi_k T^2\end{aligned} \tag{8-5-33}$$

其中　$\xi_k=\frac{1}{2}[A(X_k)f(X_k)-A(\hat{X}_k)f(\hat{X}_k)]$

再把$f(X_k)$在$\hat{X}_k$邻近展开，取一阶近似，得

$$\begin{aligned}f(X_k)&=f(\hat{X}_k)+\frac{\partial f(X)}{\partial X}\bigg|_{X=\hat{X}_k}(X_k-\hat{X}_k)\\&=f(\hat{X}_k)+A(\hat{X}_k)(X_k-\hat{X}_k)\end{aligned} \tag{8-5-34}$$

代入$\widetilde{X}_{k+1/k}$表达式中，得

$$X_{k+1}-\hat{X}_{k+1/k}=[1+A(\hat{X}_k)T](X_k-\hat{X}_k)+\Gamma_k W_k+\zeta_k T^2 \tag{8-5-35}$$

记 $$\Phi_{k+1/k}=1+A(\hat{X}_k)T \tag{8-5-36}$$

因此得预报误差方差阵为

$$P_{k+1/k}=\Phi_{k+1/k}P_k\Phi_{k+1/k}^T+Q_k^* \tag{8-5-37}$$

其中 $$Q_k^*=\Gamma_k Q_k\Gamma_k^T+T^4E[\zeta_k\zeta_k^T]$$

Q_k^*中既包含了模型噪声的方差，又包含了线性化误差的方差，它随着T增大而增大。

再考虑连续时间非线性测量方程

$$Z(t)=h(X(t))+V(t) \tag{8-5-38}$$

经采样就可得到离散时间非线性测量方程

$$Z_{k+1}=h(X_{k+1})+V_{k+1} \tag{8-5-39}$$

进而对它进行线性化。在$\hat{X}_{k+1/k}$点邻近按泰勒级数展开，取一阶近似，得

$$Z_{k+1}=h(\hat{X}_{k+1/k})+H_{k+1}(X_{k+1}-\hat{X}_{k+1/k})+\eta_{k+1}+V_{k+1} \tag{8-5-40}$$

式中 $$H_{k+1}=\left.\frac{\partial h(X)}{\partial X}\right|_{X=\hat{X}_{k+1/k}} \tag{8-5-41}$$

η_{k+1}为线性化误差

得 $$Z_{k+1}=H_{k+1}X_{k+1}+h(\hat{X}_{k+1/k})-H_{k+1}\hat{X}_{k+1/k}+\eta_{k+1}+V_{k+1}$$

记 $$y_{k+1}=h(\hat{X}_{k+1/k})-H_{k+1}\hat{X}_{k+1/k} \tag{8-5-42}$$

则 $$Z_{k+1}=H_{k+1}X_{k+1}+y_{k+1}+\eta_{k+1}+V_{k+1} \tag{8-5-43}$$

综上所述，得离散化模型的滤波递推公式如下：

预报估计。为使预报较准，可取

$$\hat{X}_{k+1/k}=\hat{X}_k+f(\hat{X}_k)T+A(\hat{X}_k)f(\hat{X}_k)\frac{T^2}{2} \tag{8-5-44}$$

式中 $$A(\hat{X}_k)=\left.\frac{\partial f(X)}{\partial X}\right|_{X=\hat{X}_k} \tag{8-5-45}$$

预报误差方差阵

$$P_{k+1/k}=\Phi_{k+1/k}P_k\Phi_{k+1/k}^T+Q_k^* \tag{8-5-46}$$

式中 $$\Phi_{k+1/k}=1+A(\hat{X}_k)T \tag{8-5-47}$$

Q_k^*包括原状态噪声方差与状态方程线性化误差方差。

增益阵

$$K_{k+1}=P_{k+1/k}H_{k+1}^T[H_{k+1}P_{k+1/k}H_{k+1}^T+R_{k+1}^*] \tag{8-5-48}$$

式中 $$H_{k+1}=\left.\frac{\partial h(X)}{\partial X}\right|_{X=\hat{X}_{k+1/k}} \tag{8-5-49}$$

R_{k+1}^*中包含原测量噪声方差与测量方程线性化误差方差。

滤波值

$$\hat{X}_{k+1}=\hat{X}_{k+1/k}+K_{k+1}[Z_{k+1}-h(\hat{X}_{k+1/k})] \tag{8-5-50}$$

滤波误差方差阵

$$P_{k+1}=(1-K_{k+1}H_{k+1})P_{k+1/k}(1-K_{k+1}H_{k+1})^T+K_{k+1}R_{k+1}^*K_{k+1}^T \tag{8-5-51}$$

§8-6 连续时间卡尔曼滤波

考察下列连续时间信号模型

$$\dot{X}(t)=A(t)X(t)+W(t) \tag{8-6-1}$$

$$Z(t)=H(t)X(t)+V(t) \tag{8-6-2}$$

其中$W(t)$和$V(t)$为互不相关的零均值白噪声过程，其协方差阵分别为

$$\operatorname{cov}[W(t),\ W(\tau)]=E[W(t)W^T(\tau)]=Q(t)\delta(t-\tau) \tag{8-6-3}$$

$$\operatorname{cov}[V(t),\ V(\tau)]=E[V(t)V^T(\tau)]=R(t)\delta(t-\tau) \tag{8-6-4}$$

$\delta(t)$为狄拉克函数，$\delta(t)$定义为

$$\delta(t)=\begin{cases}0 & t\neq 0\\ \infty & t=0\end{cases}$$

及
$$\int_{-\infty}^{\infty}\delta(t)dt=1$$

若给定观测$Z(\tau)$，$0\leqslant\tau\leqslant t$，求$X(t)$的最优估计。

我们采用把采样无限加密，令$\Delta t\to 0$，取极限的方法来推导连续时间卡尔曼滤波器。这种方法在数学上是不够严格的，但比较直观。

第一步，导出连续时间模型与离散时间模型的关系。

对于连续时间模型（8-6-1）、（8-6-2）式，采用离散化方法得出对应的离散时间模型为

$$X_k=\Phi_{k/k-1}X_{k-1}+W_{k-1} \tag{8-6-5}$$

$$Z_k=H_kX_k+V_k \tag{8-6-6}$$

其中
$$\left.\begin{aligned}&X(t)=X_{k-1},\quad X(t+\Delta t)=X_k\\&I+A(t)\Delta t=\Phi_{k/k-1}\\&Z(t+\Delta t)=Z_k,\quad H(t+\Delta t)=H_k\\&W(t)\Delta t=W_{k-1},\quad V(t+\Delta t)=V_k\end{aligned}\right\} \tag{8-6-7}$$

对于(8-6-7)式，需要作一些解释：

对于状态方程

$$\dot{X}(t)=AX(t)+W(t)\qquad X(t)|_{t=t_0}=X(t_0)$$

求得其解为

$$X(t)=e^{A(t-t_0)}X(t_0)+\int_{t_0}^{t}e^{A(t-\tau)}W(\tau)d\tau,\quad t_0\leqslant\tau\leqslant t \tag{8-6-8}$$

作变量置换，令$t_0\to t$，$t\to t+\Delta t$，则有

$$X(t+\Delta t)=e^{A\Delta t}X(t)+\int_{t}^{t+\Delta t}e^{A(t+\Delta t-\tau)}W(\tau)d\tau,\quad t\leqslant\tau\leqslant t+\Delta t \tag{8-6-9}$$

再令 $e^{A\Delta t}=\Phi$

$$t\to(k-1)T,\quad t+\Delta t\to kT$$

则有
$$X(kT)=\Phi X[(k-1)T]+\int_{t}^{t+\Delta t}e^{A(t+\Delta t-\tau)}W(\tau)d\tau \tag{8-6-10}$$

当 Δt 很小时，有

$$\Phi = e^{A\Delta t} \approx 1 + A\Delta t \tag{8-6-11}$$

记

$$\int_t^{t+\Delta t} e^{A(t+\Delta t-\tau)} W(\tau) d\tau = W[(k-1)T] \tag{8-6-12}$$

而

$$\int_t^{t+\Delta t} e^{A(t+\Delta t-\tau)} W(\tau) d\tau$$

$$\approx W(t) - A^{-1} e^{(t+\Delta t-\tau)} \Big|_t^{t+\Delta t}$$

$$\approx W(t) - A^{-1}(1 - e^{A\Delta t})$$

$$\approx W(t) \cdot \Delta t \tag{8-6-13}$$

由此得

$$W_{k-1} = W(t) \cdot \Delta t \tag{8-6-14}$$

相应的离散化状态方程为

$$X_k = \Phi X_{k-1} + W_{k-1} \tag{8-6-15}$$

推广到系数矩阵 $A(t)$ 为时变情况，则相应的离散化状态方程应为

$$X_k = \Phi_{k/k-1} X_{k-1} + W_{k-1}$$

状态噪声 W_{k-1} 的协方差可表达为

$$Q_{k-1} = E[W_{k-1} W_{k-1}^T]$$

$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t} E[W(t) W^T(t)] \Delta t^2$$

$$= \lim_{\Delta t \to 0} Q(t) \Delta t \tag{8-6-16}$$

对于测量方程

$$Z(t) = H(t) X(t) + V(t)$$

令 $t \to t + \Delta t$，则

$$Z(t+\Delta t) = H(t+\Delta t) X(t+\Delta t) + V(t+\Delta t) \tag{8-6-17}$$

对应的离散化测量方程为

$$Z_k = H_k X_k + V_k$$

V_k 的协方差为

$$R_k = E[V_k V_k^T] = \lim_{\Delta t \to 0} \frac{1}{\Delta t} E[V(t+\Delta t) V^T(t+\Delta t)] \tag{8-6-18}$$

或写成

$$\lim_{\Delta t \to 0} R_k \Delta t = \lim_{\Delta t \to 0} E[V(t+\Delta t) V^T(t+\Delta t)] = R(t) \tag{8-6-19}$$

第二步，通过对离散时间滤波器取极限得到连续时间滤波器。

已知离散时间卡尔曼滤波器的滤波方程为

$$\hat{X}_k = \Phi_{k/k-1} \hat{X}_{k-1} + K_k [Z_k - H_k \Phi_{k/k-1} \hat{X}_{k-1}] \tag{8-6-20}$$

利用(8-6-7)式，得

$$\hat{X}(t+\Delta t) = [1 + A(t)\Delta t] \hat{X}(t)$$

$$+ K_k(t+\Delta t)[Z(t+\Delta t) - H(t+\Delta t)(1 + A(t)\Delta t) \hat{X}(t)] \tag{8-6-21}$$

稍加整理，写成

$$\frac{\hat{X}(t+\Delta t)-\hat{X}(t)}{\Delta t}=A(t)\hat{X}(t)+\frac{K_k(t+\Delta t)}{\Delta t}[Z(t+\Delta t)$$

$$-H(t+\Delta t)(\hat{X}(t)-A(t)\hat{X}(t)\Delta t)] \qquad (8\text{-}6\text{-}22)$$

令 $\Delta t \to 0$，对(8-6-22)式取极限，并令

$$\lim_{\Delta t\to 0}\frac{K_k(t+\Delta t)}{\Delta t}=K(t) \qquad (8\text{-}6\text{-}23)$$

则得

$$\dot{\hat{X}}(t)=A(t)\hat{X}(t)+K(t)[Z(t)-H(t)\hat{X}(t)] \qquad (8\text{-}6\text{-}24)$$

又已知离散时间增益方程为

$$K_k=P_{k/k-1}H_k^T[H_kP_{k/k-1}H_k^T+R_k]^{-1} \qquad (8\text{-}6\text{-}25)$$

两边除以 Δt，再利用(8-6-7)式，得

$$\frac{K_k(t+\Delta t)}{\Delta t}=\frac{1}{\Delta t}P(t+\Delta t/t)H(t+\Delta t)^T$$

$$\times[H(t+\Delta t)P(t+\Delta t/t)H(t+\Delta t)^T+R_k]^{-1}$$

$$=P(t+\Delta t/t)H(t+\Delta t)^T[H(t+\Delta t)$$

$$\times P(t+\Delta t/t)H(t+\Delta t)^T\Delta t+R_k\Delta t]^{-1} \qquad (8\text{-}6\text{-}26)$$

两边取极限，得

$$\lim_{\Delta t\to 0}\frac{K_k(t+\Delta t)}{\Delta t}=K(t) \qquad (8\text{-}6\text{-}27)$$

$$\lim_{\Delta t\to 0}P(t+\Delta t/t)$$

$$=\lim_{\Delta t\to 0}[(1+A(t)\Delta t)P(t/t)(1+A(t)\Delta t)+Q(t)\Delta t]$$

$$=P(t/t) \qquad (8\text{-}6\text{-}28)$$

于是得连续时间的增益矩阵为

$$K(t)=P(t/t)H^T(t)R^{-1}(t) \qquad (8\text{-}6\text{-}29)$$

再由离散时间状态方程

$$P_k=(1-K_kH_k)P_{k/k-1} \qquad (8\text{-}6\text{-}30)$$

得

$$P(t+\Delta t/t+\Delta t)$$

$$=P(t+\Delta t/t)-K(t+\Delta t)H(t+\Delta t)P(t+\Delta t/t) \qquad (8\text{-}6\text{-}31)$$

其中

$$P(t+\Delta t/t)=[1+A(t)\Delta t]P(t/t)[1+A(t)\Delta t]^T+Q_{k-1}$$

$$\approx P(t/t)+[A(t)P(t/t)+P(t/t)A^T(t)]\Delta t+Q_{k-1} \qquad (8\text{-}6\text{-}32)$$

代入(8-6-30)式中，得

$$P(t+\Delta t/t+\Delta t)=P(t/t)+[A(t)P(t/t)+P(t/t)A^T(t)]\Delta t$$

$$+Q_{k-1}-K(t+\Delta t)H(t+\Delta t)P(t+\Delta t/t) \qquad (8\text{-}6\text{-}33)$$

令 $\Delta t \to 0$，取极限得

$$\lim_{\Delta t\to 0}\frac{P(t+\Delta t/t+\Delta t)-P(t/t)}{\Delta t}$$

$$=\lim_{\Delta t\to 0}[A(t)P(t/t)+P(t/t)A^T(t)]$$

$$+\lim_{\Delta t\to 0}\left[\frac{1}{\Delta t}Q_{k-1}-\frac{1}{\Delta t}K(t+\Delta t)H(t+\Delta t)P(t+\Delta t/t)\right] \qquad (8\text{-}6\text{-}34)$$

得连续时间滤波误差协方差阵为

$$\dot{P}(t/t)=A(t)P(t/t)+P(t/t)A^T(t)+Q(t)$$
$$-P(t/t)H^T(t)R^{-1}(t)H(t)P(t/t) \tag{8-6-35}$$

总结一下，连续时间卡尔曼滤波器方程为

滤波估计

$$\dot{\hat{X}}(t)=A(t)\hat{X}(t)+K(t)[Z(t)-H(t)\hat{X}(t)] \tag{8-6-36}$$

最佳增益

$$K(t)=P(t/t)H^T(t)R^{-1}(t) \tag{8-6-37}$$

误差协方差阵

$$\dot{P}(t/t)=A(t)P(t/t)+P(t/t)A^T(t)+Q(t)$$
$$-P(t/t)H^T(t)R^{-1}(t)H(t)P(t/t) \tag{8-6-38}$$

初始条件应当是

$$\hat{X}(0)=\bar{X}(0) \qquad P(0)=\mathrm{var}\,X(0) \tag{8-6-39}$$

连续时间的信号模型和卡尔曼滤波器框图如图8-6-1和图8-6-2所示。

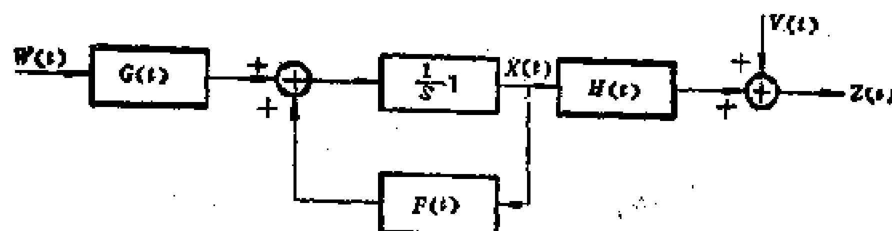


图8-6-1　连续时间信号模型

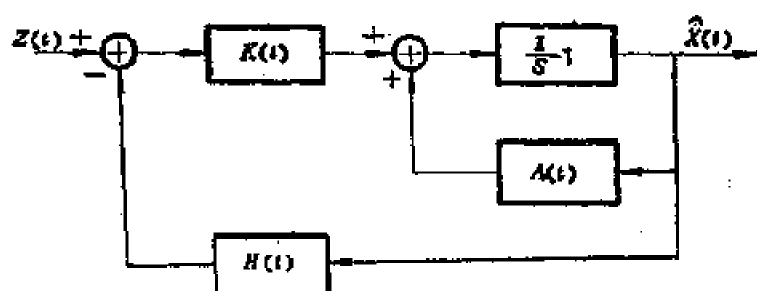


图8-6-2　连续时间卡尔曼滤波器

对于定常信号模型，有

$$\left.\begin{aligned}\dot{X}(t)&=AX(t)+GW(t)\\Z(t)&=HX(t)+V(t)\end{aligned}\right\} \tag{8-6-40}$$

其中A、G、H均是常值矩阵，$W(t)$、$V(t)$为不相关的零均值平稳　噪声过程

$$E[W(t)]=0 \qquad E[V(t)]=0$$
$$E[W(t+\tau)W^T(t)]=Q\delta(\tau)$$
$$E[V(t+\tau)V^T(t)]=R\delta(\tau)$$

类似于离散时间的卡尔曼滤波器，在模型满足可观测性与可控制性条件下，时间充分长之后，不管初值选取如何，滤波器进入稳态，方差阵$P(t/t)$趋于常值，故有$\dot{P}(t/t)=0$。

稳态滤波误差阵可由下式解出

$$F\overline{P}+\overline{P}F^{T}-\overline{P}H^{T}R^{-1}H\overline{P}+GQG^{T}=0 \tag{8-6-41}$$

稳态增益为

$$\overline{K}=\overline{P}HR^{-1} \tag{8-6-42}$$

从而状态估计 $\hat{X}(t)$ 满足下列方程

$$\dot{\hat{X}}(t)=A\hat{X}(t)+\overline{K}[Z(t)-H\hat{X}(t)] \tag{8-6-43}$$

例8-6-1 设有连续时间模型

$$\dot{X}(t)=\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}X(t)+\begin{pmatrix}0\\ 1\end{pmatrix}W(t)$$

$$Z(t)=[1,\ 0]X(t)+V(t)$$

$$Q(t)=\sigma_w^2 \qquad R(t)=\sigma_v^2$$

解：记 $\overline{P}=\begin{pmatrix}p_{11} & p_{12}\\ p_{21} & p_{22}\end{pmatrix}$

代入(8-4-31)式，得

$$\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}\begin{pmatrix}p_{11} & p_{12}\\ p_{21} & p_{22}\end{pmatrix}+\begin{pmatrix}p_{11} & p_{12}\\ p_{21} & p_{22}\end{pmatrix}\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}$$

$$-\begin{pmatrix}p_{11} & p_{12}\\ p_{21} & p_{22}\end{pmatrix}\begin{pmatrix}1\\ 0\end{pmatrix}\frac{1}{\sigma_v^2}[1,0]\begin{pmatrix}p_{11} & p_{12}\\ p_{21} & p_{22}\end{pmatrix}+\begin{pmatrix}0\\ 1\end{pmatrix}\sigma_w^2[1,0]=0$$

因为协方差矩阵的 $p_{12}=p_{21}$，展开后，得到下面三个方程：

$$\left.\begin{aligned}&2p_{12}-\frac{1}{\sigma_v^2}p_{11}^2=0\\ &p_{22}-\frac{1}{\sigma_v^2}p_{11}p_{12}=0\\ &-\frac{1}{\sigma_v^2}p_{12}^2+\sigma_w^2=0\end{aligned}\right\} \tag{8-6-44}$$

为使 $\overline{P}$ 保持正定，一律取正值解，得

$$\left.\begin{aligned}&p_{11}=\sigma_v\sqrt{2\sigma_w\sigma_v}\\ &p_{12}=\sigma_w\sigma_v\\ &p_{22}=\sigma_w\sqrt{\sigma_w\sigma_v}\end{aligned}\right\} \tag{8-6-45}$$

稳态增益

$$\overline{K}=\overline{P}H^{T}R^{-1}=\begin{pmatrix}\sqrt{\dfrac{2\sigma_w}{\sigma_v}}\\ \dfrac{\sigma_w}{\sigma_v}\end{pmatrix} \tag{8-6-46}$$

稳态滤波方程为

$$\dot{\hat{X}}(t)=\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}\hat{X}(t)+\begin{pmatrix}\sqrt{\dfrac{2\sigma_w}{\sigma_v}}\\ \dfrac{\sigma_w}{\sigma_v}\end{pmatrix}[Z(t)-[1,0]\hat{X}(t)] \qquad (8\text{-}6\text{-}47)$$

解此微分方程，即可求得状态估计 $\hat{X}(t)$。

一般来说，连续时间卡尔曼滤波器由于计算困难而不直接应用，主要用来分析研究，以便得出连续时间信号的最优估计的可能方案。

例8-6-2　假设发送信号是由一阶微分方程

$$\dot{X}(t)=a x(t)+w(t), \qquad t>0$$

所描述的随机过程的样本函数，已知

$$a>0, \qquad E[x(0)]=0, \qquad E[x^2(0)]=\sigma_0^2$$

$w(t)$为零均值白噪声，方差为$E[w(t)w(\tau)]=\sigma_1^2\delta(t-\tau)$。

假设传送信号是

$$z(t)=x(t)+v(t)$$

其中 $v(t)$ 是零均值白噪声，方差为$E[v(t)v(t)]=\sigma_2^2(t-\tau)$，又 $x(0)$、$w(t)$、$v(t)$三者互相独立。

现在的问题要确定接收机的特性，使收到的噪声最少，即要求滤波误差

$$\tilde{x}(t)=x(t)-\hat{X}(t)$$

的方差最小。

解：由题意已知$A(t)=-a$，　$H(t)=1$，　$p_0=\sigma_0^2$，　$Q(t)=\sigma_1^2$，　$R(t)=\sigma_2^2$ 代入方差方程，得

$$\dot{p}=-2ap-\frac{1}{\sigma_2^2}p^2+\sigma_1^2 \qquad p(0)=\sigma_0^2$$

为求解此微分方程，采用分离变量法得

$$\frac{dp}{p^2+2a\sigma_2^2p-\sigma_1^2\sigma_2^2}=-\frac{dt}{\sigma_2^2}$$

记方程式　　$p^2+2a\sigma_2^2p-\sigma_1^2\sigma_2^2=0$

的两根为 p_1 和 p_2，由此得

$$\frac{dp}{(p-p_1)(p-p_2)}=\frac{1}{p_1-p_2}\left[\frac{dp}{p-p_1}-\frac{dp}{p-p_2}\right]=-\frac{dz}{\sigma_2^2}$$

因为　　$p_1-p_2=2\sigma_2^2\sqrt{a^2+\dfrac{\sigma_1^2}{\sigma_2^2}}$

所以得

$$\frac{dp}{p-p_1}-\frac{dp}{p-p_2}=\left(-2\sqrt{a^2+\frac{\sigma_1^2}{\sigma_2^2}}\right)dz$$

对上式两端积分，稍加整理，并记

$$\mu=\sqrt{a^2+\frac{\sigma_1^2}{\sigma_2^2}}$$

就得到

$$\frac{p(t)-\rho_1}{p(t)-\rho_2}=ce^{-2\mu t}$$

其中 c 是积分常数，代入初始条件 $p(0)=\sigma_0^2$，可解出

$$c=\frac{\sigma_0^2-\rho_1}{\sigma_0^2-\rho_2}$$

由此得出估计误差的方差方程为

$$p(t)=\frac{\rho_1-\rho_2 c e^{-2\mu t}}{1-c e^{-2\mu t}}$$

增益方程为

$$k(t)=\frac{p(t)}{\sigma_2^2}$$

滤波方程为

$$\dot{\hat{x}}(t)=-a\hat{x}(t)+\frac{p(t)}{\sigma_2^2}\left[z(t)-\hat{x}(t)\right]$$

当 $t\rightarrow\infty$ 时，就得到稳态滤波器如下：

$$p(t)\rightarrow\rho_1=\left[-a+\sqrt{a^2+\frac{\sigma_1^2}{\sigma_2^2}}\right]\sigma_2^2$$

$$k(t)\rightarrow\frac{\rho_1}{\sigma_2^2}=-a+\sqrt{a^2+\frac{\sigma_1^2}{\sigma_2^2}}$$

因此在稳态情况下，最佳接收机是一个常系数线性系统。

$$\dot{\hat{x}}(t)=-a\hat{x}(t)+\frac{\rho_1}{\sigma_2^2}z(t)+\left[a-\sqrt{a^2+\frac{\sigma_1^2}{\sigma_2^2}}\right]\hat{x}(t)$$

或

$$\dot{\hat{x}}(t)=-\mu\hat{x}(t)+\frac{\rho_1}{\sigma_2^2}z(t)$$

根据上式就可构造出时不变最佳稳态滤波器，如图 8-6-3 所示，其中

$$\mu=\left(a^2+\frac{\sigma_1^2}{\sigma_2^2}\right)^{1/2}$$

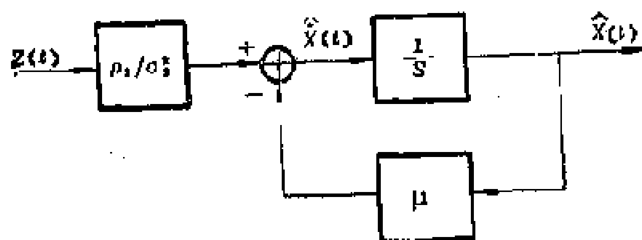


图8-6-3　连续时间稳态滤波器

第九章 卡尔曼滤波的应用

在第七和第八章中，已详细研究了卡尔曼滤波的性质与算法，本章将讨论卡尔曼滤波的工程应用问题。由于卡尔曼滤波的的应用十分广泛，本书中不可能一一涉及，本章只对卡尔曼滤波在声纳，雷达中的应用进行较详细的研究。具体将研究如何根据实际情况来构造信号模型和相应的滤波器，为方便起见，假定噪声统计量已用其他方法获得。对于需要实时估计噪声统计量的情况留在下章研究。

§9-1 应用卡尔曼滤波时应注意的问题

前两章已经介绍了卡尔曼滤波的基本思想和算法，正确掌握这些理论结果无疑是十分重要的。但是，我们应当认识到，仅仅掌握这些理论结果，尚不足以保证有效地应用卡尔曼滤波，实践证明，应用中必须注意下面几个问题。

1. 数学模型的建立

卡尔曼滤波的出发点是动态过程的状态方程和测量方程，即数学模型，其中比较难以确定的是状态方程。建立的模型要反映对象的基本特征，又要足够简化，以保证计算的实现。数学模型的好坏，常常是滤波成败的关键。

状态方程可通过力学的、电学的以及其他各种物理定律建立。首先应选择坐标系和状态变量，状态变量的多少应能表现动态过程的主要特征，建立的模型尽可能保证是线性的，噪声尽可能保证是白色的，这时滤波器的计算最简便，精度最好。还应保证建立的模型是可观测的和可控制的，对于不可观测的和不可控制的信号模型，是不可能构造卡尔曼滤波器的。

然而在许多场合下，不能以基本物理定律来建立状态方程，这时就要采用试验法。对于一系统施加输入信号，并观测输出信号，然后通过对试验数据的分析来得到动态过程的描述。一般来说，建立模型要经过确定模型结构、估计模型参数和模型验证三个步骤。上述课题是现代控制理论的一个分支——模型识别的中心内容。

2. 验前统计量

验前统计量包括初值和噪声方差，基于卡尔曼滤波器收敛性的研究，对于初值的确定无需过分注意，至于状态噪声方差阵Q_k和测量噪声方差阵R_k的选取，由于它们的取值直接影响到增益K_k的值，对滤波效果有很大影响，所以应给予较多注意。在实际工作中，装定的P_0、Q_k、R_k应当比真值稍大些，容易保证滤波器稳定。倘若对噪声统计值一无所知，则需要采用下章中的自适应方法。

3. 计算问题

卡尔曼滤波进行递推计算的每一步，都有方差矩阵的求逆计算。理论上协方差阵是对称的和正定的，但在有限字长计算机上计算时，由于截尾的影响，使协方差阵失去对称性和正定性，从而造成滤波器发散，对这个问题在具体应用时，亦不应忽略。

上述问题中，最难以解决的是关于噪声统计值。在卡尔曼滤波早期应用中，通常采用大量实验方法或用其他物理方法来获得先验知识，但是困难较大且很难得到精确估计。当噪声统计值为时变时尤其困难，自七十年代以来，自适应方法得到重视，目前仍在发展之中。

卡尔曼滤波的应用十分广泛，这里仅选择若干个例子加以考察。这里的方案并不一定是最佳的，而仅是可能的方案，并且为了方便起见，在本章讨论中暂时假设对噪声的统计值已有先验知识。

§9-2　在航行管理雷达中的应用

航行管理雷达大都采用边扫描边跟踪体制，它具有等速旋转的天线，天线每旋转一周，能给出一次目标位置数据，斜距r_k和方位角β_k，某些雷达还可提供目标仰角ε_k。由于目标常常有多个，数据率又比较低，为了跟踪并描绘出每个目标的航迹，需要应用卡尔曼滤波器。航行管理雷达跟踪目标的流程如图9-2-1所示。

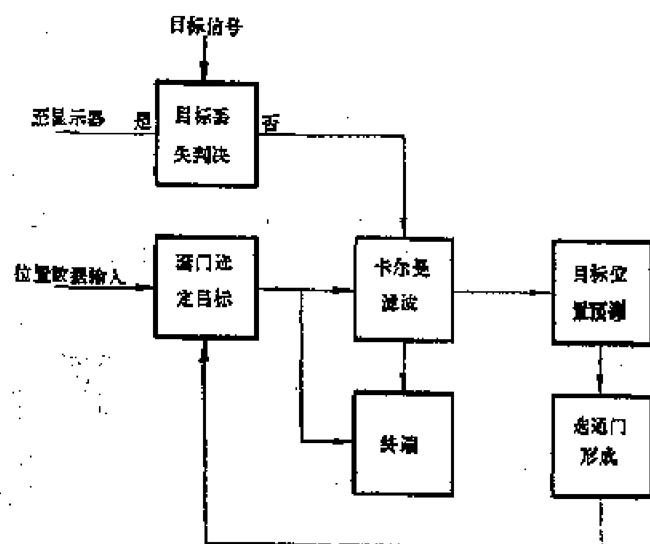


图9-2-1　航管雷达目标航迹显示

假设在t_k时刻送入某一目标位置数据Z_k，通过卡尔曼滤波器计算出目标当前位置的估计以及对t_{k+1}时刻位置的预报，此预报量在选通门形成器中形成适当大小的选通门，送至套门与目标选择器，当天线旋转一周后，得到该目标在t_{k+1}时刻上的位置数据Z_{k+1}，此时位置选通门也正好位于Z_{k+1}，将套住的该目标位置数据Z_{k+1}再送入卡尔曼滤波器中，进行下一轮计算。从卡尔曼滤波器得到的位置估计$\hat{Z}_k$可以直接提供应用，如送入显示终端，也可以通过航迹平滑器使航迹更光滑。使用多个选通门，可以跟踪多个目标并分别描绘出航迹。对于丢失的目标，通过丢失判定，即撤消对该目标的跟踪。该系统已成功地应用于空中交通管理系统，简称为航管雷达。现在详细考察航管雷达中的卡尔曼滤波器的设计问题。

方案一，建立数学模型的第一步是选择坐标系和确定状态变量。因为测量是在极坐标系中进行的，测量噪声n_r和n_β可以看成是白噪声，因此选定极坐标系作为建立数学模型的基础。考虑到目标基本上以等型直线航行，但换算到极坐标系中，可以认为其速度扰动为时间相关过程。假定速度扰动为一阶马尔科夫过程，其相关函数为

$$\Phi_k=\sigma_u^2\rho^{|k|},\quad \rho<1 \tag{9-2-1}$$

根据§8-4的论述，采用有色状态噪声时的算法，把有色速度扰动u_k表达成由白噪声w_k驱动一阶递归滤波器所产生，即有

$$u_{k+1}=\rho u_k+w_k \tag{9-2-2}$$

若u_k的方差为σ_u^2，则w_k的方差为$\sigma_u^2(1-\rho^2)$，其中σ_u^2和ρ需要根据当时目标飞行情况作出

选择。综上所述，我们选择平面极坐标系、选择状态矢量为$X=[r,\ \dot r,\ u_r,\beta,\dot\beta,u_\beta]^T$，列出状态方程如下：

$$\begin{pmatrix} r\\ \dot r\\ u_r\\ \beta\\ \dot\beta\\ u_\beta\end{pmatrix}_{k+1}=\begin{pmatrix}1 & T & 0 & & & \\ 0 & 1 & 1 & & 0 & \\ 0 & 0 & \rho & & & \\ & & & 1 & T & 0\\ & 0 & & 0 & 1 & 1\\ & & & 0 & 0 & \rho\end{pmatrix}\begin{pmatrix} r\\ \dot r\\ u_r\\ \beta\\ \dot\beta\\ u_\beta\end{pmatrix}_k+\begin{pmatrix}0\\0\\w_r\\0\\0\\w_\beta\end{pmatrix}\tag{9-2-3}$$

写成矢量形式，记

$$\Phi=\begin{pmatrix}1 & T & 0 & & 0 & \\ 0 & 1 & 1 & & & \\ 0 & 0 & \rho & 1 & T & 0\\ & & & 0 & 1 & 1\\ & 0 & & 0 & 0 & \rho\end{pmatrix},\ \Gamma W_k=\begin{pmatrix}0\\0\\w_r\\0\\0\\w_\beta\end{pmatrix}=\begin{pmatrix}0&0\\0&0\\1&0\\0&0\\0&0\\0&1\end{pmatrix}\begin{pmatrix}w_r\\w_\beta\end{pmatrix}$$

则有

$$X_{k+1}=\Phi X_k+\Gamma W_k\tag{9-2-4}$$

再列出测量方程

$$Z_k=HX_k+V_k\tag{9-2-5}$$

其中$Z_k=\begin{bmatrix}r_k\\ \beta_k\end{bmatrix}$　　$V_k=\begin{bmatrix}n_{r,k}\\ n_{\beta,k}\end{bmatrix}$　　$H=\begin{bmatrix}1&0&0&0&0&0\\0&0&0&1&0&0\end{bmatrix}$

(9-2-4)和(9-2-5)式就构成了信号模型，并且容易验证此模型是可观测和可控制的。

第二步是确定噪声的统计知识。令σ^2_{ur}表示目标线速度扰动的方差，令$\sigma^2_{u\beta}$表示目标角速度扰动的方差。因为角速度扰动是由线速度扰动所引起，因此，k时刻的角速度扰动的方差可由线速度扰动的方差算出。

$$(\sigma^2_{u\beta})_k=\frac{(\sigma^2_{ur})_k}{r_k^2}\tag{9-2-6}$$

其中σ^2_{ur}可以根据当时的飞行情况估算出来，但r_k是待估计量，是未知值，故用预报值$\hat r_{k/k-1}$代替之，于是得出状态噪声方差阵Q_k为

$$Q_k=E[W_kW_k^T]=\begin{pmatrix}\sigma^2_{ur}(1-\rho^2) & 0\\ 0 & \sigma^2_{u\beta}(1-\rho^2)\end{pmatrix}$$

$$=\begin{pmatrix}\sigma^2_{ur}(1-\rho^2) & 0\\ 0 & \dfrac{\sigma^2_{ur}(1-\rho^2)}{\hat r^2_{k/k-1}}\end{pmatrix}\tag{9-2-7}$$

因为Q_k是时变的，与预报距离$\hat r_{k/k-1}$有关，故增益K_k必须在线计算，容易得出测量噪声协方差阵R_k为

$$R_k=\begin{pmatrix}\sigma^2_r & 0\\ 0 & \sigma^2_\beta\end{pmatrix}=R\tag{9-2-8}$$

第三步是确定滤波器的初值。根据关于滤波器收敛性的结论，初值可以任意选取，但为了缩短滤波器的进入最佳状态的过渡时间，初值应尽可能准确些。下面介绍的两点法常在工程上用来确定初值$\hat{X}_0$和P_0。

取滤波器正式运行之前，t_{-1}和t_0时刻上的测量值为

$$Z_{-1}=\begin{pmatrix}\hat{r}_{-1}\\ \hat{\beta}_{-1}\end{pmatrix},\qquad Z_0=\begin{pmatrix}\hat{r}_0\\ \hat{\beta}_0\end{pmatrix}$$

取初始状态估计$\hat{X}_0$为

$$\hat{X}_0=\begin{pmatrix}\hat{r}_0\\ \hat{\dot{r}}_0\\ 0\\ \hat{\beta}_0\\ \hat{\dot{\beta}}_0\\ 0\end{pmatrix}=\begin{pmatrix}\hat{r}_0\\ \dfrac{\hat{r}_0-\hat{r}_{-1}}{T}\\ 0\\ \hat{\beta}_0\\ \dfrac{\hat{\beta}_0-\hat{\beta}_{-1}}{T}\\ 0\end{pmatrix} \tag{9-2-9}$$

由测量方程$\hat{r}_k=r_k+n_{rk}$，$\hat{\beta}_k=\beta_k+n_{\beta k}$，可得知在$t_0$时刻的目标状态应为

$$\left.\begin{aligned}&r_0=\hat{r}_0-n_{r0}\\ &\dot{r}_0=\dot{r}_{-1}+u_{r(-1)}=\frac{r_0-r_{-1}}{T}+u_{r(-1)}\\ &u_{r0}=\rho u_{r(-1)}+w_{r(-1)}\\ &\beta_0=\hat{\beta}_0-n_{\beta 0}\\ &\dot{\beta}_0=\dot{\beta}_{-1}+u_{\beta(-1)}=\frac{\beta_0-\beta_{-1}}{T}+u_{\beta(-1)}\\ &u_{\beta 0}=\rho u_{\beta(-1)}+w_{\beta(-1)}\end{aligned}\right\} \tag{9-2-10}$$

在t_0时刻的估计误差矢量为

$$\hat{X}_0-X_0=\begin{pmatrix}n_{r0}\\ \dfrac{n_{r0}-n_{r(-1)}}{T}-u_{r(-1)}\\ -w_{r(-1)}\\ n_{\beta 0}\\ \dfrac{n_{\beta 0}-n_{\beta(-1)}}{T}-u_{\beta(-1)}\\ -w_{\beta(-1)}\end{pmatrix} \tag{9-2-11}$$

因而t_0时刻的估计误差方差阵为

$$P_0=E[(\hat{X}_0-X_0)(\hat{X}_0-X_0)^T] \tag{9-2-12}$$

得

$$P_0=\begin{pmatrix}p_{11}&p_{12}&p_{13}&&&\\ p_{21}&p_{22}&p_{23}&&0&\\ p_{31}&p_{32}&p_{33}&&&\\ &&&p_{44}&p_{45}&p_{46}\\ &0&&p_{54}&p_{55}&p_{56}\\ &&&p_{64}&p_{65}&p_{66}\end{pmatrix} \tag{9-2-13}$$

其中

$$
\left.\begin{aligned}
&p_{11}=\sigma_r^2,\quad p_{22}=\sigma_{mr}^2+\frac{2\sigma_r^2}{T^2}\\
&p_{12}=p_{21}=\frac{\sigma_r^2}{T},\quad p_{33}=\sigma_{mr}^2\\
&p_{13}=p_{31}=0,\quad p_{23}=p_{32}=\rho\sigma_{mr}^2\\
&p_{44}=\sigma_\beta^2,\quad p_{55}=\sigma_{m\beta}^2+\frac{2\sigma_\beta^2}{T^2}\\
&p_{45}=p_{54}=\frac{\sigma_\beta^2}{T},\quad p_{66}=\sigma_{m\beta}^2=\frac{\sigma_{mr}^2}{\hat{r}_1^2}\\
&p_{46}=p_{64}=0,\quad p_{56}=p_{65}=\rho\sigma_{m\beta}^2
\end{aligned}\right\}\tag{9-2-14}
$$

在取得初值$\hat{X}_0$和P_0之后，即可启动卡尔曼滤波器。

方案二，考虑到目标大部分时间内是按直线方式飞行的，所以把状态方程设置在直角坐标系中更为适宜，这时不需采用扩充状态矢量法，测量仍是在极坐标内进行的。

选定状态矢量为

$$
Z_k=[x_k,\ \dot{x}_k,\ \ddot{x}_k,\ y_k,\ \dot{y}_k,\ \ddot{y}_k]^T \tag{9-2-15}
$$

列出状态方程为

$$
\begin{pmatrix} x\\ \dot{x}\\ \ddot{x}\\ y\\ \dot{y}\\ \ddot{y} \end{pmatrix}_{k+1}=
\begin{pmatrix}
1 & T & \frac{T^2}{2} & & & \\
0 & 1 & T & & 0 & \\
0 & 0 & \rho & & & \\
 & & & 1 & T & \frac{T^2}{2}\\
 & 0 & & 0 & 1 & T\\
 & & & 0 & 0 & \rho
\end{pmatrix}
\begin{pmatrix} x\\ \dot{x}\\ \ddot{x}\\ y\\ \dot{y}\\ \ddot{y} \end{pmatrix}_k+
\begin{pmatrix}
0 & 0\\ 0 & 0\\ 1 & 0\\ 0 & 0\\ 0 & 0\\ 0 & 1
\end{pmatrix}
\begin{pmatrix} w_x\\ w_y \end{pmatrix}_k \tag{9-2-16}
$$

或写成

$$
X_{k+1}=\Phi X_k+\Gamma W_k \tag{9-2-17}
$$

其中w_x、w_y为加速度扰动噪声，并假设它们是零均值白噪声，方差等于σ^2，故$Q_k=\begin{bmatrix}\sigma^2 & 0\\ 0 & \sigma^2\end{bmatrix}$。

这里ρ称为加速度相关系数，$0<\rho<1$，其物理意义是指目标速度不可能无限增长。

因为测量是在极坐标中进行的，测量值是斜距与方位角，可列出测量方程如下：

$$
\begin{pmatrix} z_{rk}\\ z_{\beta k} \end{pmatrix}=\begin{pmatrix} r_k\\ \beta_k \end{pmatrix}+\begin{pmatrix} v_{rk}\\ v_{\beta k} \end{pmatrix} \tag{9-2-18}
$$

测量噪声v_{rk}、$v_{\beta k}$可看成是零均值的噪声，其方差阵为

$$
R_k=\begin{pmatrix} \sigma_r^2 & 0\\ 0 & \sigma_\beta^2 \end{pmatrix}
$$

因为

$$
r_k=(x_k^2+y_k^2)^{\frac{1}{2}}=f_1(X_k) \tag{9-2-19}
$$

$$
\beta_k=\frac{x_k}{y_k}=f_2(X_k) \tag{9-2-20}
$$

显然，测量是状态矢量的非线性函数，所以，需要采用第八章中的广义卡尔曼滤波方法。

记
$$F(\boldsymbol{X})=\begin{pmatrix} f_1(\boldsymbol{X}) \\ f_2(\boldsymbol{X}) \end{pmatrix} \tag{9-2-21}$$

则有
$$\boldsymbol{H}_k=\frac{\partial F}{\partial \boldsymbol{X}}\bigg|_{X=\hat{X}_{k/k-1}}$$
$$=\begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial \dot{x}} & \frac{\partial f_1}{\partial \ddot{x}} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial \dot{y}} & \frac{\partial f_1}{\partial \ddot{y}} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial \dot{x}} & \frac{\partial f_2}{\partial \ddot{x}} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial \dot{y}} & \frac{\partial f_2}{\partial \ddot{y}} \end{pmatrix}_{X=\hat{X}_{k/k-1}}$$
$$=\begin{pmatrix} \frac{x}{r} & 0 & 0 & \frac{y}{r} & 0 & 0 \\ \frac{y}{r^2} & 0 & 0 & -\frac{x}{r^2} & 0 & 0 \end{pmatrix}_{X=\hat{X}_{k/k-1}} \tag{9-2-22}$$

其中$r_k=\sqrt{x_k^2+y_k^2}$，于是得到线性化测量方程为
$$\boldsymbol{Z}_{k+1}=\boldsymbol{H}_{k+1}\boldsymbol{X}_{k+1}+\boldsymbol{Y}_{k+1}+\boldsymbol{V}_{k+1} \tag{9-2-23}$$

其中
$$\boldsymbol{H}_{k+1}=\frac{\partial F}{\partial \boldsymbol{X}}\bigg|_{X=\hat{X}_{k+1/k}} \tag{9-2-24}$$
$$\boldsymbol{Y}_{k+1}=F(\hat{\boldsymbol{X}}_{k+1/k})-\boldsymbol{H}_{k+1}\hat{\boldsymbol{X}}_{k+1/k} \tag{9-2-25}$$

所以相应的广义卡尔曼滤波器为
$$\hat{\boldsymbol{X}}_{k+1/k}=\Phi\hat{\boldsymbol{X}}_k \tag{9-2-26}$$
$$\boldsymbol{P}_{k+1/k}=\Phi\boldsymbol{P}_k\Phi^T+\boldsymbol{Q}_k \tag{9-2-27}$$
$$\boldsymbol{K}_{k+1}=\boldsymbol{P}_{k+1/k}\boldsymbol{H}_{k+1}^T[\boldsymbol{H}_{k+1}\boldsymbol{P}_{k+1/k}\boldsymbol{H}_{k+1}^T+\boldsymbol{R}^*]^{-1} \tag{9-2-28}$$
$$\hat{\boldsymbol{X}}_{k+1}=\hat{\boldsymbol{X}}_{k+1/k}+\boldsymbol{K}_{k+1}[\boldsymbol{Z}_{k+1}-F(\hat{\boldsymbol{X}}_{k+1/k})] \tag{9-2-29}$$
$$\begin{aligned}\boldsymbol{P}_{k+1}&=(1-\boldsymbol{K}_{k+1}\boldsymbol{H}_{k+1})\boldsymbol{P}_{k+1/k}(1-\boldsymbol{K}_{k+1}\boldsymbol{H}_{k+1})\\&+\boldsymbol{K}_{k+1}\boldsymbol{R}_{k+1}^*\boldsymbol{K}_{k+1}^T\end{aligned} \tag{9-2-30}$$

由于状态矢量设置在直角坐标系中，故初值$\hat{X}_0$和P_0也应设置在直角坐标系中，为了使初值较准确，也可采用两点法确定之，这时需要把测量值$(\hat{r}_{-1},\hat{\beta}_{-1})$、$(\hat{r}_0,\hat{\beta}_0)$换算到直角坐标系中。

方案三，把模型设置在平面直角坐标系中，选择状态矢量为
$$\boldsymbol{X}_k=[x_k,\ \dot{x}_k,\ \ddot{x}_k,\ y_k,\ \dot{y}_k,\ \ddot{y}_k]^T \tag{9-2-31}$$

列出状态方程为
$$\begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \\ y \\ \dot{y} \\ \ddot{y} \end{pmatrix}_{k+1}=\begin{pmatrix} 1 & T & \frac{T^2}{2} & & & \\ 0 & 1 & T & & 0 & \\ 0 & 0 & \rho & & & \\ & & & 1 & T & \frac{T^2}{2} \\ & 0 & & 0 & 1 & T \\ & & & 0 & 0 & \rho \end{pmatrix}\begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \\ y \\ \dot{y} \\ \ddot{y} \end{pmatrix}_k+\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} w_x \\ w_y \end{pmatrix}_k \tag{9-2-32}$$

其中ρ为加速度相关系数，$0<\rho<1$；w_x、w_y为加速度扰动噪声。假定它们是零均值白噪声，方差相等为σ^2，则得出状态噪声方差阵$\boldsymbol{Q}_k=\boldsymbol{Q}=\begin{bmatrix}\sigma^2 & 0\\ 0 & \sigma^2\end{bmatrix}$ (9-2-33)

测量是在极坐标内进行的，测量值是斜距r_k和方位角θ_k。已知测量噪声的方差为σ_r^2和σ_θ^2，现在的问题是要在直角坐标系中列出测量方程。

设测量误差为e_r和e_θ，由几何关系可得出

$$\left.\begin{aligned}e_x&=e_r\sin\theta+re_\theta\cos\theta\\ e_y&=e_r\cos\theta-re_\theta\sin\theta\end{aligned}\right\}\tag{9-2-34}$$

记

$$\boldsymbol{V}_k=\begin{pmatrix}e_{xk}\\ e_{yk}\end{pmatrix}=\begin{pmatrix}e_r\sin\theta+re_\theta\cos\theta\\ e_r\cos\theta-re_\theta\cos\theta\end{pmatrix}_k\tag{9-2-35}$$

因为e_r、e_θ为零均值白噪声，互相独立，故有

$$E[\boldsymbol{V}_k]=0$$

$$\boldsymbol{R}_k=E[\boldsymbol{V}_k,\boldsymbol{V}_k^T]$$

$$=\begin{pmatrix}\sigma_r^2\sin^2\theta_k+r_k^2\sigma_\theta^2\cos^2\theta_k, & (\sigma_r^2-r_k^2\sigma_\theta^2)\sin\theta_k\cos\theta_k\\ (\sigma_r^2-r_k^2\sigma_\theta^2)\sin\theta_k\cos\theta_k, & \sigma_r^2\cos^2\theta_k+r_k^2\sigma_\theta^2\sin^2\theta_k\end{pmatrix}\tag{9-2-36}$$

因为上式中r_k、θ_k均为待估计量，故用估计值$\hat{r}_{k/k-1}$和$\hat{\theta}_{k/k-1}$代替之。

于是可以列出平面直角坐标系中的测量方程如下：

$$\begin{pmatrix}r_k\sin\theta_k\\ r_k\cos\theta_k\end{pmatrix}=\begin{pmatrix}1&0&0&0&0&0\\0&0&0&1&0&0\end{pmatrix}\begin{pmatrix}x\\ \dot{x}\\ \ddot{x}\\ y\\ \dot{y}\\ \ddot{y}\end{pmatrix}+\begin{pmatrix}e_{xk}\\ e_{yk}\end{pmatrix}\tag{9-2-37}$$

由以上讨论可见，对同一运动目标，可以选取不同的模型，构造出不同的滤波器。滤波效果的好坏将主要决定于模型的精确性与噪声参数的精确性。此外，还应考虑计算量的大小，不要使模型过于复杂。

三维空间内的卡尔曼滤波器

假设雷达在三维空间内跟踪一个大体上按等速直线飞行的目标，要求设计跟踪卡尔曼滤波器。

我们选择直角坐标系，并选择状态矢量

$$\boldsymbol{X}=[x,\dot{x},y,\dot{y},z,\dot{z}]\tag{9-2-38}$$

而雷达测量数据是在球坐标系中有斜距r、方位角β和仰角ε，如图9-1-2所示。

坐标系之间的关系为

$$r=(x^2+y^2+z^2)^{1/2}=f_1(x,y,z)\tag{9-2-39}$$

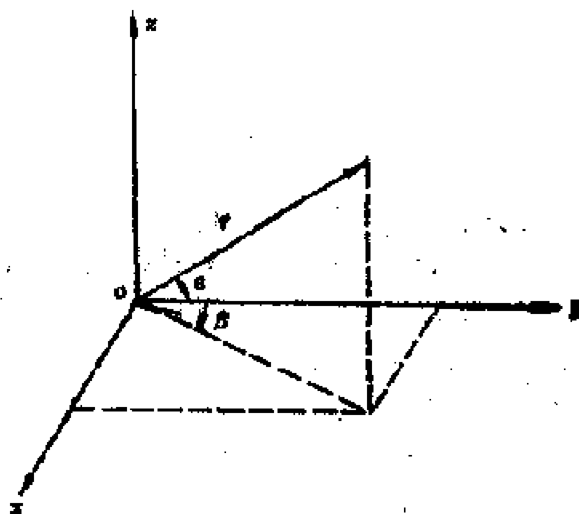


图9-1-2 球坐标系与直角坐标系

$$\beta = \tan^{-1}\frac{x}{y} = f_2(x, y, z) \tag{9-2-40}$$

$$\varepsilon = \tan^{-1}\frac{z}{(x^2 + y^2)^{\frac{1}{2}}} = f_3(x, y, z) \tag{9-2-41}$$

故测量方程可表达为

$$\begin{pmatrix} r \\ \beta \\ \varepsilon \end{pmatrix}_k = \begin{pmatrix} f_1(x, y, z) \\ f_2(x, y, z) \\ f_3(x, y, z) \end{pmatrix}_k + \begin{pmatrix} v_r \\ v_\beta \\ v_\varepsilon \end{pmatrix}_k \tag{9-2-42}$$

变换至直角坐标系后的线性化测量方程为

$$H_k = \left.\frac{\partial F}{\partial X}\right|_{X = \hat{X}_{k/k-1}}$$

$$= \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial \dot{x}} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial \dot{y}} & \frac{\partial f_1}{\partial z} & \frac{\partial f_1}{\partial \dot{z}} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial \dot{x}} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial \dot{y}} & \frac{\partial f_2}{\partial z} & \frac{\partial f_2}{\partial \dot{z}} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial \dot{x}} & \frac{\partial f_3}{\partial y} & \frac{\partial f_3}{\partial \dot{y}} & \frac{\partial f_3}{\partial z} & \frac{\partial f_3}{\partial \dot{z}} \end{pmatrix}_{X = \hat{X}_{k/k-1}}$$

$$= \begin{pmatrix} \frac{x}{r} & 0 & \frac{y}{r} & 0 & \frac{z}{r} & 0 \\ \frac{y}{x^2 + y^2} & 0 & \frac{-x}{x^2 + y^2} & 0 & 0 & 0 \\ \frac{-xz}{r^2\sqrt{x^2 + y^2}} & 0 & \frac{-yz}{r^2\sqrt{x^2 + y^2}} & 0 & \frac{\sqrt{x^2 + y^2}}{r^2} & 0 \end{pmatrix}_{X = \hat{X}_{k/k-1}} \tag{9-2-43}$$

相应的状态方程为

$$\begin{pmatrix} x \\ \dot{x} \\ y \\ \dot{y} \\ z \\ \dot{z} \end{pmatrix}_k = \begin{pmatrix} 1 & T & & & & \\ 0 & 1 & & & & \\ & & 1 & T & & \\ & & 0 & 1 & & \\ & & & & 1 & T \\ & & & & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \\ y \\ \dot{y} \\ z \\ \dot{z} \end{pmatrix}_{k-1} + \begin{pmatrix} 0 \\ w_x \\ 0 \\ w_y \\ 0 \\ w_z \end{pmatrix}_{k-1} \tag{9-2-44}$$

其中w_x、w_y、w_z为三个坐标方向上的速度扰动噪声。

然后应用广义卡尔曼滤波器就可得出目标状态矢量的估计。

另外，我们也可以选择在球坐标系中设置状态方程。在三维空间中目标作直线运动时，若用球坐标(r, β, ε)来描述目标的运动，需要用相当复杂的非线性方程，现设法用一个简单的方法对目标运动方程作线性近似，这样就可以在球坐标系内进行线性递推滤波。

运动模型线性化的目的是在球坐标系中得到一个近似线性模型，下面的展开式可以用来达到这个目的。

设在直角坐标系中的状态矢量为

$$X_k=[x_k, \dot{x}_k, y_k, \dot{y}_k, z_k, \dot{z}_k]^T \tag{9-2-45}$$

当采样间隔较小时，可以把r_{k+1}以r_k为原点展开成泰勒级数，并取其一次近似，则得出

$$r_{k+1}=r_k+\frac{\partial r}{\partial X}\bigg|_k(x_{k+1}-x_k)+\frac{\partial r}{\partial y}(y_{k+1}-y_k)+\frac{\partial r}{\partial z}\bigg|_k(z_{k+1}-z_k)+\text{高阶项} \tag{9-2-46}$$

若目标作直线运动并含有加速度扰动噪声w_x、w_y和w_z，则可得

$$\left.\begin{aligned} x_{k+1}-x_k&=\dot{x}_kT+w_{xk}\frac{T^2}{2}\\ y_{k+1}-y_k&=\dot{y}_kT+w_{yk}\frac{T^2}{2}\\ z_{k-1}-z_k&=\dot{z}_kT+w_{zk}\frac{T^2}{2}\end{aligned}\right\} \tag{9-2-47}$$

此外，可求出

$$\frac{\partial r}{\partial x}=\frac{x}{r},\quad \frac{\partial r}{\partial y}=\frac{y}{r},\quad \frac{\partial r}{\partial z}=\frac{z}{r} \tag{9-2-48}$$

把以上关系代入可得

$$r_{k+1}=r_k+\frac{x_k}{r_k}\left(\dot{x}_kT+w_{xk}\frac{T^2}{2}\right)+\frac{y_k}{r_k}\left(\dot{y}_kT+w_{yk}\frac{T^2}{2}\right)+\frac{z_k}{r_k}\left(\dot{z}_kT+w_{zk}\frac{T^2}{2}\right) \tag{9-2-49}$$

因为
$$\dot{r}_k=\frac{x_k\dot{x}_k}{r_k}+\frac{y_k\dot{y}_k}{r_k}+\frac{z_k\dot{z}_k}{r_k} \tag{9-2-50}$$

故得
$$r_{k-1}=r_k+\dot{r}_kT+\frac{T^2}{2r_k}[x_kw_{xk}+y_kw_{yk}+z_kw_{zk}] \tag{9-2-51}$$

把方向余弦用球坐标来表达，则有

$$\begin{aligned}\frac{x_k}{r_k}&=\cos(x_k, r_k)=\sin\beta_k\cos\varepsilon_k\\ \frac{y_k}{r_k}&=\cos(y_k, r_k)=\cos\beta_k\cos\varepsilon_k\\ \frac{z_k}{r_k}&=\cos(z_k, r_k)=\sin\varepsilon_k\end{aligned} \tag{9-2-52}$$

于是，可以得出在球坐标系中目标在斜距上的运动方程为

$$\begin{pmatrix} r\\ \dot{r}_k\end{pmatrix}_{k+1}=\begin{pmatrix}1 & T\\ 1 & 0\end{pmatrix}\begin{pmatrix} r\\ \dot{r}\end{pmatrix}_k$$

$$
+\begin{pmatrix} \frac{T^2}{2}\sin\beta_k\cos\varepsilon_k & \frac{T^2}{2}\cos\beta_k\cos\varepsilon_k & \frac{T^2}{2}\sin\varepsilon_k \\ T\sin\beta_k\cos\varepsilon_k & T\cos\beta_k\cos\varepsilon_k & T\sin\varepsilon_k \end{pmatrix}\begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix}_k \tag{9-2-53}
$$

同理可得出目标在方位角上的运动方程为

$$
\begin{pmatrix} \beta \\ \dot{\beta} \end{pmatrix}_{k+1} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \beta \\ \dot{\beta} \end{pmatrix}_k
+\begin{pmatrix} \frac{T^2\cos\beta_k}{2r_k\cos\varepsilon_k} & -\frac{T^2\sin\beta_k}{2r_k\cos\varepsilon_k} & 0 \\ \frac{T\cos\beta_k}{r_k\cos\varepsilon_k} & -\frac{T\sin\beta_k}{r_k\cos\varepsilon_k} & 0 \end{pmatrix}\begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix}_k \tag{9-2-54}
$$

目标在俯仰角上的运动方程为

$$
\begin{pmatrix} \varepsilon \\ \dot{\varepsilon} \end{pmatrix}_{k+1} = \begin{pmatrix} 1 & T \\ 1 & 0 \end{pmatrix}\begin{pmatrix} \varepsilon \\ \dot{\varepsilon} \end{pmatrix}_k
+\begin{pmatrix} \frac{-T^2\sin\beta_k\sin\varepsilon_k}{2r_k} & \frac{-T^2\cos\beta_k\sin\varepsilon_k}{2r_k} & \frac{T^2\cos\varepsilon_k}{2r_k} \\ \frac{-T\sin\beta_k\sin\varepsilon_k}{r_k} & \frac{-T\cos\beta_k\sin\varepsilon_k}{r_k} & \frac{T\cos\varepsilon_k}{r_k} \end{pmatrix}\begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix}_k \tag{9-2-55}
$$

合并以上三式，可得球坐标系中的目标的状态方程为

$$
X_{k+1} = \Phi X_k + \Gamma_k W_k \tag{9-2-56}
$$

其中 $X_{k+1} = [r,\ \dot{r},\ \beta,\ \dot{\beta},\ \varepsilon,\ \dot{\varepsilon}]^T_{k+1}$

$W_k = [w_x, w_y, w_z]^T_k$

$$
\Phi = \begin{pmatrix} 1 & T & & & & \\ 0 & 1 & & & 0 & \\ & & 1 & T & & \\ & & 0 & 1 & & \\ & 0 & & & 1 & T \\ & & & & 0 & 1 \end{pmatrix}
$$

$$
\Gamma_k = \begin{pmatrix} \frac{T^2}{2}\sin\beta_k\cos\varepsilon_k & \frac{T^2}{2}\cos\beta_k\cos\varepsilon_k & \frac{T^2}{2}\sin\varepsilon_k \\ T\sin\beta_k\cos\varepsilon_k & T\cos\beta_k\cos\varepsilon_k & T\sin\varepsilon_k \\ \frac{T^2\cos\beta_k}{2r_k\cos\varepsilon_k} & \frac{-T^2\sin\beta_k}{2r_k\cos\varepsilon_k} & 0 \\ \frac{T\cos\beta_k}{r_k\cos\varepsilon_k} & \frac{-T\sin\beta_k}{r_k\cos\varepsilon_k} & 0 \\ \frac{-T^2\sin\beta_k\sin\varepsilon_k}{2r_k} & \frac{-T^2\cos\beta_k\sin\varepsilon_k}{2r_k} & \frac{T^2\cos\varepsilon_k}{2r_k} \\ \frac{-T\sin\beta_k\sin\varepsilon_k}{r_k} & \frac{-T\cos\beta_k\sin\varepsilon_k}{r_k} & \frac{T\cos\varepsilon_k}{r_k} \end{pmatrix}
$$

在球坐标系中的测量方程为

$$\begin{pmatrix} r \\ \beta \\ \varepsilon \end{pmatrix}_{k+1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} r \\ \dot{r} \\ \beta \\ \dot{\beta} \\ \varepsilon \\ \dot{\varepsilon} \end{pmatrix}_{k+1} + \begin{pmatrix} v_r \\ v_\beta \\ v_\varepsilon \end{pmatrix}_{k+1} \tag{9-2-57}$$

根据以上模型，就可构造相应的滤波器。

§9-3 在被动声纳中的应用

主动声纳的工作原理与雷达相似，但数据率很低，故卡尔曼滤波的应用有一定困难，本节着重介绍卡尔曼滤波在被动声纳中的应用。

所谓被动声纳，就是声纳本身不发射信号，只接收目标舰船发出的噪声，这类噪声是宽带的，通过水下信道传播到声纳接收水听器。由于水声信道对高频的强烈衰减作用，距离稍远时，只有大约千赫以下的频率才能传播到水听器而被接收。为了对声源目标定位，被动声纳的水听器通常布成某种形式的阵列，被动声纳对声源目标的定位与跟踪通常采用下列两种方法。

一、三元阵系统

设有三个水听器构成一线列阵，如图9-3-1所示。水听器1、2、3分别接收来自目标舰船发出的连续噪声，用相关器可以测得信号到达的时间差τ_{12}和τ_{23}。

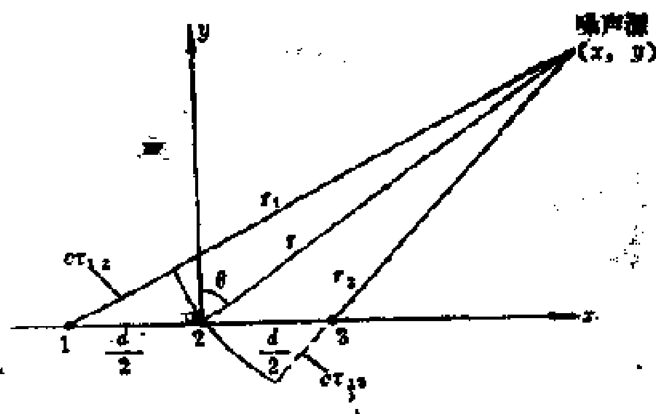


图9-3-1 三元线列阵对噪声源定位

设噪声源以柱面波方式传播，由几何关系得

$$c\tau_{12} = r_1 - r = \sqrt{r^2 + \left(\frac{d}{2}\right)^2 + rd\sin\theta} - r$$

$$= r\left[\sqrt{1 + \frac{1}{r^2}\left(\frac{d}{2}\right)^2 + \frac{d}{r}\sin\theta} - 1\right] \tag{9-3-1}$$

因为 $\qquad \frac{1}{r^2}\left(\frac{d}{2}\right)^2 \ll 1, \quad \frac{d}{r}\sin\theta \ll 1$

令 $\qquad x = \frac{d^2}{4r^2} + \frac{d}{r}\sin\theta$

由 $\qquad \sqrt{1+x} = 1 + \frac{x}{2} - \frac{x^2}{8} + \cdots \tag{9-3-2}$

经近似化简，得 $\quad c\tau_{12} \approx \frac{d}{2}\sin\theta + \frac{d^2}{8r}\cos^2\theta \tag{9-3-3}$

同理，得 $$c\tau_{23} \approx \frac{d}{2}\sin\theta - \frac{d^2}{8r}\cos^2\theta \tag{9-3-4}$$

令 $\tau_C = \tau_{12} + \tau_{23}$，$\tau_D = \tau_{12} - \tau_{23}$

则由(9-3-3)、(9-3-4)式，得

方位角 $$\theta = \sin^{-1}\frac{c\tau_x}{d} \tag{9-3-5}$$

斜距 $$r = \frac{d^2\cos^2\theta}{4c\tau_D} \tag{9-3-6}$$

其中 c 为声在水中传播速度。

设 $\delta\tau_{12}$、$\delta\tau_{23}$ 为测量误差，包括相关器的测时误差和波阵面的随机弯曲等因素，假定它们是零均值互相独立的白噪声，方差等于

$$E[(\delta\tau_{12})^2] = E[(\delta\tau_{23})^2] = \sigma^2$$

则测向误差为

$$\delta\theta = \frac{c}{d\cos\theta}\delta\tau_x = \frac{c}{d\cos\theta}(\delta\tau_{12} + \delta\tau_{23}) \tag{9-3-7}$$

其方差为

$$\sigma_\theta^2 = E[\delta_\theta^2] = \left(\frac{c}{d\cos\theta}\right)^2(2\sigma^2) \tag{9-3-8}$$

测距误差为

$$\begin{aligned}\delta r &= -\frac{d^2\cos^2\theta}{4c\tau_D^2}\delta\tau_D \\ &= -\left(\frac{d^2\cos^2\theta}{4c}\right)\left(\frac{4cr}{d^2\cos^2\theta}\right)^2\delta\tau_D \\ &= -\frac{4cr^2}{d^2\cos^2\theta}\delta\tau_D\end{aligned} \tag{9-3-9}$$

其方差为

$$\sigma_r^2 = E[(\delta r)^2] = \left(\frac{4cr^2}{d^2\cos^2\theta}\right)^2(2\sigma^2) \tag{9-3-10}$$

由此可得出三元阵法对声源定位的不确定椭圆，其长轴半径和短轴半径分别为

$$\sigma_r = \frac{4\sqrt{2}\,cr^2}{d^2\cos^2\theta}\sigma$$

$$r\sigma_\theta = \frac{\sqrt{2}\,cr}{d\cos\theta}\sigma$$

可见，σ_r与r^2 成比例，$r\sigma_\theta$ 与 r 成比例，并且不确定椭圆的形状随声源的空间位置而变，如图 9-3-2 所示。所以这种方法的测距误差是非常大的，或者说，其定位能力是很低的，甚至是不可信的。

上面说过测时误差方差，σ^2之中包含相关器本身的测量误差和波阵面的随机起伏，所以仅仅改进相关器的测量精度并不能显著改善三元阵法的定位能力，因而有必要再通过卡尔曼滤波器来提高其定位能力。

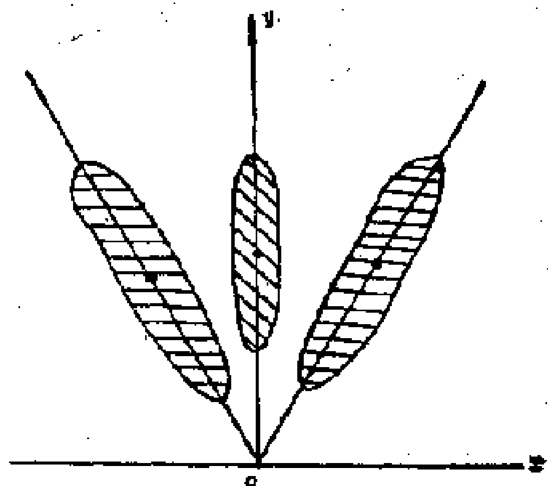
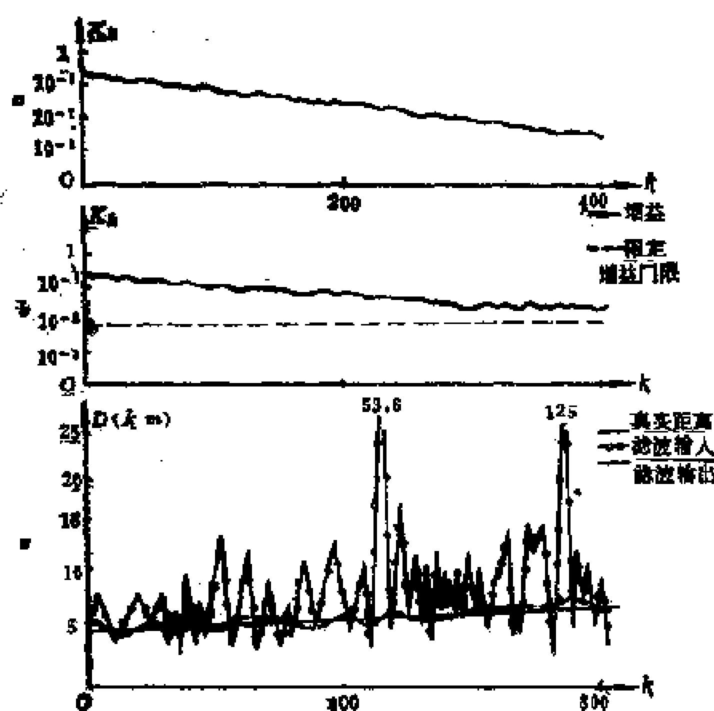


图9-3-2　三元阵定位时的不确定椭圆

某被动声纳站中的卡尔曼滤波

法国生产的某型被动声纳，采用的是三元阵法测距，在该声纳站中，为了提高测距精度，在互相关器的设计方面采取了一些措施。它把 2～10kHz的宽带噪声用滤波器分成四路窄带噪声，然后对每个通道用狭带相关器测量时延，再对四个时延值求加权和作为总的时延的估计，其中应用了内插法来改善互相关器的时延测量精度。



该型声纳装在潜艇上时，三元阵的总长是 46m，阵元间距为 23m。图9-3-3c为该型声纳在未采用后置卡尔曼滤波时的海上实测结果，这些数据的实验条件是：目标在本艇的正横方向，以 2～7 节的速度从4.2km处背离本艇驶去，得到的距离竟在1.4～125km 范围内剧烈起伏（见图9-3-3c）。

图 9-3-3 给出了该声纳的测距性能，图 a 为未加限定增益时卡尔曼滤波器中距离增益的变

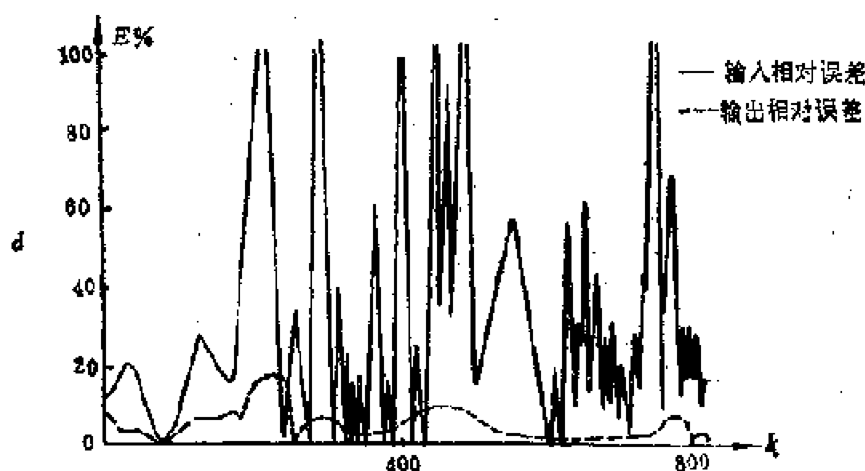


图9-3-3　某被动声纳的测距性能

化；图 b 为加上限定增益时卡尔曼滤波器中的距离增益变化；图 c 为目标真实距离，卡尔曼滤波器输入端的距离估计，卡尔曼滤波器输出端的距离估计三者的比较；图 d 为相对测距误差的比较。

该被动声纳采取的后置处理包括一个一阶低通自回归数字滤波器和一个二阶卡尔曼滤波器，设计者声称，该声纳在海上试验中测距精度为距离的10%左右。

下面再详细研究该型声纳中卡尔曼滤波器的设计。首先列出信号模型。

由(9-3-6)式，得 k 时刻时延差 τ_D 为

$$\tau_{Dk}=\frac{d^2\cos^2\theta_k}{4c}\cdot\frac{1}{r_k}=b(\theta_k)\cdot\frac{1}{r_k}$$

τ_{Dk} 的测量值为

$$\hat{\tau}_{Dk}=b(\theta_k)\cdot\frac{1}{r_k}+v_k \tag{9-3-11}$$

其中 v_k 表示量测噪声，包括相关器测时误差与波阵面的随机起伏。

记
$$u_k=\frac{1}{r_k},\quad \dot{u}_k=\frac{d}{dt}\left(\frac{1}{r}\right)_k=-\frac{1}{r_k^2}\left(\frac{dr}{dt}\right)_k$$

于是，可写出测量方程为

$$\hat{\tau}_{Dk}=[b(\theta_k),\ 0]\begin{pmatrix}u_k\\ \dot{u}_k\end{pmatrix}+v_k \tag{9-3-12}$$

再列出状态方程，用 r_k、$\dot{r}_k$ 表示目标在 k 时刻的距离和速度，假设目标以等速度航行，且没有动态噪声，则运动方程为

$$r_k=r_{k-1}+T\dot{r}_{k-1}$$

取倒数，得

$$\frac{1}{r_k}=\frac{1}{r_{k-1}+T\dot{r}_{k-1}}=\frac{r_{k-1}-T\dot{r}_{k-1}}{r_{k-1}^2-(T\dot{r}_{k-1})^2}$$

利用关系式 $r_{k-1}\gg T\dot{r}_{k-1}$，得

$$\frac{1}{r_k}=\frac{1}{r_{k-1}}-T\frac{\dot{r}_{k-1}}{r_{k-1}^2} \tag{9-3-13}$$

因为
$$\frac{d}{dt}\left(\frac{1}{r}\right)_{k-1}=-\frac{1}{r_{k-1}^2}\frac{d}{dt}r_{k-1}=-\frac{\dot{r}_{k-1}}{r_{k-1}^2} \tag{9-3-14}$$

令 $u_k=\dfrac{1}{r_k}$，$u_{k-1}=\dfrac{1}{r_{k-1}}$，$\dfrac{d}{dt}\left(\dfrac{1}{r_{k-1}}\right)=\dot{u}_{k-1}$，由(9-3-13)式，得

$$u_k=u_{k-1}+T\dot{u}_{k-1}$$

又因为目标是没有加速度的，$\dot{r}_k=\dot{r}_{k-1}$，故有

$$\dot{u}_k\approx\dot{u}_{k-1}$$

于是可以写出矢量状态方程为

$$\begin{pmatrix}u_k\\ \dot{u}_k\end{pmatrix}=\begin{pmatrix}1 & T\\ 0 & 1\end{pmatrix}\begin{pmatrix}u_{k-1}\\ \dot{u}_{k-1}\end{pmatrix} \tag{9-3-15}$$

(9-3-12)与(9-3-15)构成了简化的信号模型，根据此模型即可构造卡尔曼滤波器，其中

$$b(\theta_k)=\frac{d^2\cos^2\theta_k}{4c},\quad \theta_k=\sin^{-1}\frac{c}{d}(\tau_{12k}+\tau_{23k})$$

由上述信号模型即可构造卡尔曼滤波器。由于没有状态噪声，卡尔曼滤波器的增益是递降的，一旦目标有机动时，最终必将导致滤波器的发散。为了防止发散，又采用限定增益的办法，即限定滤波器增益不得小于某个门限，关于限定增益以防止滤波器发散的原理将在下章中研究。

显见，该声纳站所采用的卡尔曼滤波器是次优的，可以断言，它有很大的改进余地。

二、四元阵系统

上述三元阵法曾被某些被动声纳所采用，实践结果表明，尤其是在浅海，直达声、海底反射、海面反射等多途干涉的结果，使得两个分开的水听器接收波形之间有较大的不相似性，称之为去相关现象。

图 9-3-4 是根据爆炸声测量的信号水平横向相关性，水文条件是等强层。图中 d 表示两个水听器之间的距离，显然随着的增大，信号的水平空间相关性下降。

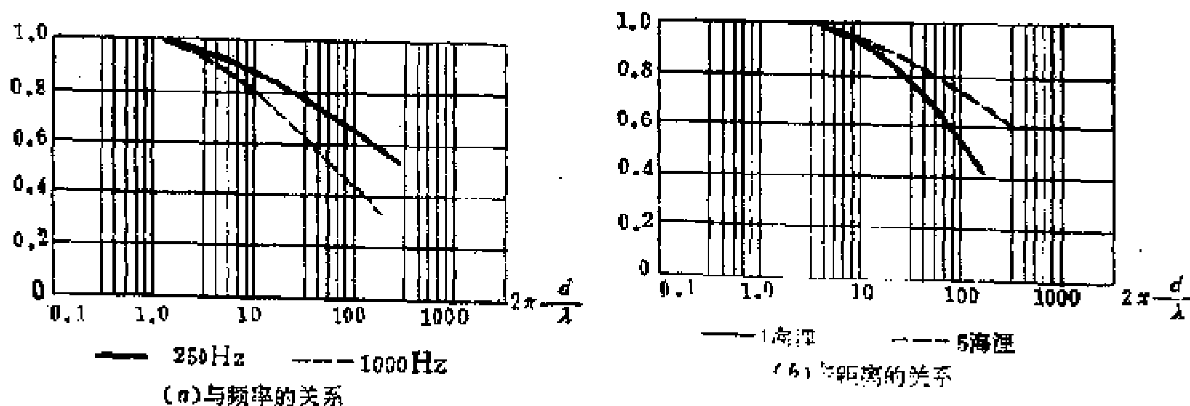


图9-3-4　水平横向空间相关性（英国北海，冬季，水深65m）

由图 a 还可看出，水平横向相关性随信号频率的增加而下降；由图 b 可看出，声源距离远时，水平横向相关性增大。

上述实验结果表明，利用加大基元间距方法是不能提高定位能力的，甚至适得其反。改进相关器的设计也不能显著提高被动声纳站的定位能力。因此提出了下述四元阵被动定位方法。

四元阵定位原理如下：阵元1、2和阵元3、4各构成一对测向系统，每个子阵的间距为 l，l 限在10m以下，以保持良好的相关性，但两个子阵的间距可以适当大，四元阵定位原理见图9-3-5。

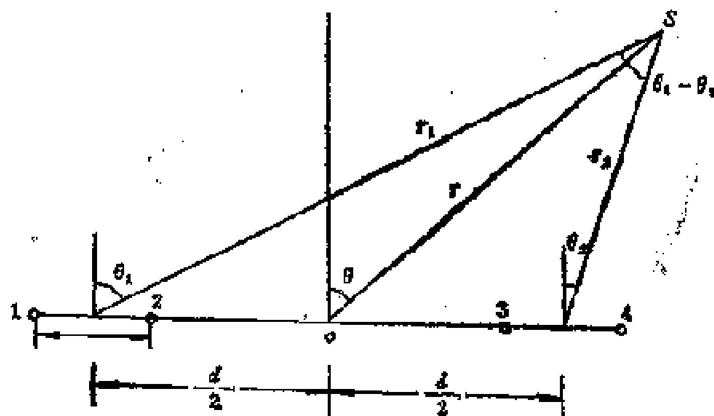


图9-3-5　四元阵对噪声源定位原理

由相关器 1 测得时差τ_{12}，得

$$\theta_1 = \sin^{-1}\frac{c\tau_{12}}{l}$$

由相关器 2 测得时差τ_{34}，得

$$\theta_2 = \sin^{-1}\frac{c\tau_{34}}{l}$$

令时间差 $\tau_D = \tau_{12} - \tau_{34}$，

$$\tau_D = \frac{l}{c}(\sin\theta_1 - \sin\theta_2)$$

$$= \frac{2l}{c}\cos\frac{\theta_1+\theta_2}{2}\sin\frac{\theta_1-\theta_2}{2} \tag{9-3-16}$$

一般情况下有 $r \gg d$，$\sin\frac{\theta_1-\theta_2}{2} \approx \frac{\theta_1-\theta_2}{2}$ (9-3-17)

和 $\theta_1 - \theta_2 \approx \frac{d\cos\theta}{r}$，$\theta \approx \frac{\theta_1+\theta_2}{2}$ (9-3-18)

代入(9-3-16)式，得

目标距离 $$r = \frac{ld}{c}\cdot\frac{\cos^2\theta}{\tau_D} \tag{9-3-19}$$

目标方位 $$\theta = \frac{\theta_1+\theta_2}{2} \tag{9-3-20}$$

再来计算不确定椭圆，相关器测时误差对距离测量的影响为

$$\delta_r = -\frac{d\,l\cos^2\theta}{c\tau_D^2}\delta\tau_D$$

$$= -\frac{cr^2}{d\,l\cos^2\theta}\delta\tau_D \tag{9-3-21}$$

设两个相关器的测时误差为零均值白噪声，方差等于σ^2，互相独立，则测距误差的方差为

$$\sigma_r^2 = E[(\delta r)^2] = \frac{c^2 r^4}{d^2 l^2\cos^4\theta}(2\sigma^2) \tag{9-3-22}$$

又相关器测时误差对方位测量误差的影响为

$$\delta\theta = \frac{1}{2}(\delta\theta_1 + \delta\theta_2) = \frac{c}{2l\cos\theta}(\delta\tau_{12} + \delta\tau_{34}) \tag{9-3-23}$$

其方差为

$$\sigma_\theta^2 = E[(\delta\theta)^2] = \frac{c^2}{4l^2\cos^2\theta}(2\sigma^2) \tag{9-3-24}$$

可见定位不确定椭圆的长轴和短轴半径分别为

$$\sigma_r = \frac{\sqrt{2}\,cr^2}{d\,l\cos^2\theta}\sigma \tag{9-3-25}$$

$$r\sigma_\theta = \frac{1}{\sqrt{2}}\frac{cr}{l\cos\theta}\sigma \tag{9-3-26}$$

二次相关器的平稳效应

测距的关键在于提取距离信号τ_D，τ_D用二次相关器提取，二次相关器的处理原理如图9-3-6所示。

波阵面的起伏可看成是叠加在方位角θ上的噪声，用$\delta\theta$表示之，$\delta\theta$与$\delta\tau_{12}$和$\delta\tau_{34}$的关系为

$$\delta\tau_{12} \approx \frac{l}{c}\cos\theta_1\,\delta\theta, \quad \delta\tau_{34} \approx \frac{l}{c}\cos\theta_2\,\delta\theta \tag{9-3-27}$$

$\delta\theta$与$\delta\tau_D$的关系为

$$\delta\tau_D \approx -\frac{2ld}{cr}\cos\theta\sin\theta\,\delta\theta \tag{9-3-28}$$

因为$\frac{d}{r}\ll 1$，可见$\delta\theta$对τ_{12}、τ_{34}的影响很大，而对τ_D的影响很小，因此可以把τ_{12}、τ_{34}看成是快变化的非平稳过程，而把τ_D看成是慢变化的准平稳过程，故一次相关器的积分时间只能取得较短，而二次相关器的积分时间可比一次相关器增大一至二个数量级，从而提高了信噪比，并得到一个比较稳定的相关函数波形。

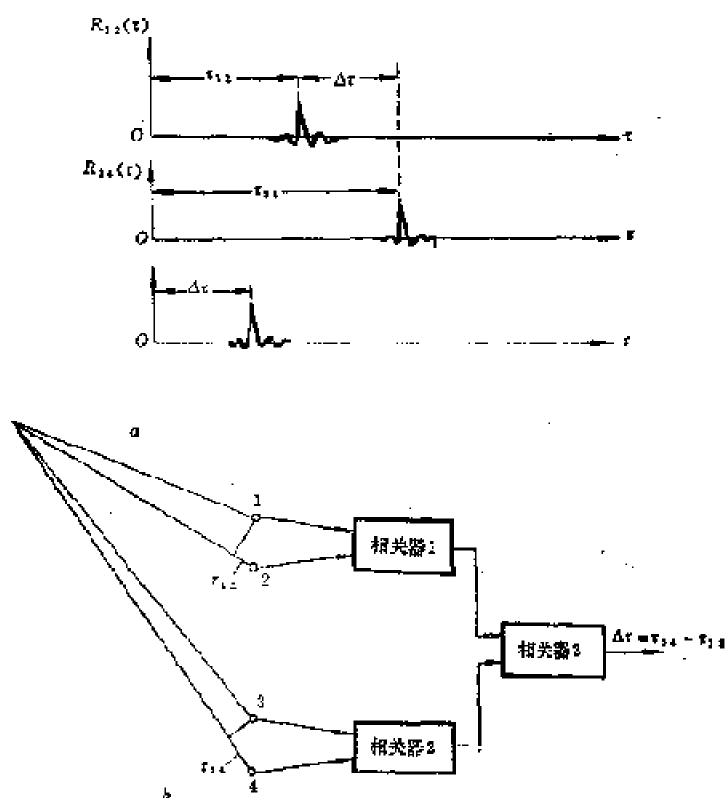


图9-3-6　四元阵二次相关处理原理

四元阵与三元阵方法比较，允许使用较大的线阵总长，提高了测距精度，并且由于采用了二次相关器，可以抑制波阵面起伏的影响，但在设备上略有增加。在这两种定位方法中，都引入了距离远大于阵长的近似，故所得定位公式只适用于远距离目标的位置测量。

上面我们讨论了噪声源的被动方式定位问题，也就是一次测量定位。对于静止声源，可以采用多次测量，然后用最小二乘估计法以得出较好的位置估计；对于运动中目标，我们必须保留τ_{12}、τ_{34}、τ_D中的运动信息，抑制波阵面随机起伏和相关器的测量噪声，故需要对τ_{12}、τ_{34}、τ_D采用某种过滤措施，即数据平滑处理。

数据平滑处理

海上试验证明，由于海洋声信道起伏的影响，时差τ_{12}、τ_{34}以及τ_D并非严格遵循几何关系而呈现随机起伏。随机起伏有快速起伏和长周期慢起伏两类，快速起伏经过一次相关

和二次相关得到了抑制，而长周期的慢起伏引起τ_D的随机漂移，在恶劣水文条件下，τ_D可偏离准确值达数百微秒，所以，欲实现稳定而准确的测距，必须克服这种长周期随机漂移因素的影响，恢复τ_{12}、τ_{34}、τ_D的几何规律。为此，应对测量数据τ_{12}、τ_{34}、τ_D进行必要的平滑处理。

对于被动测距声纳来说，不论采用何种方案，测距误差都很大，采用平滑处理是十分必要的，后置平滑的方法有很多，例如α-β滤波，卡尔曼滤波等。理论上说卡尔曼滤波是线性无偏最小方差估计，但在目前情况下，由于缺乏噪声先验统计知识，不易得到好的效果，这里采用了比较简易的线性拟合法。

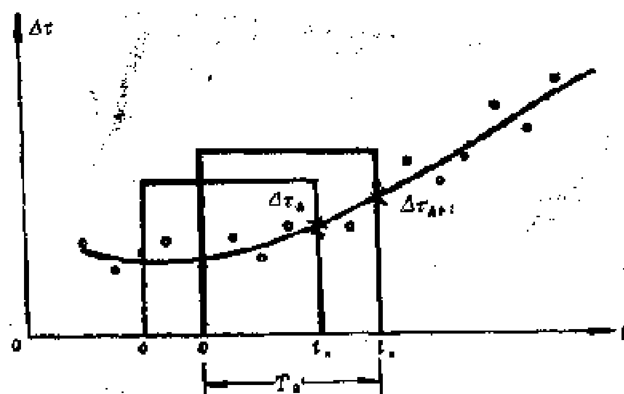


图9-3-7 滑窗式线性拟合，$\Delta\tau_k$的平滑

以τ_D的平滑处理为例，设$\tau_D=\Delta\tau$随时间t的变化关系如图9-3-7所示。今欲估计t_k时刻的$\Delta\tau_k$，设在滑动时间窗T_0内共录取了n采样值$\Delta\tau_1,\cdots,\Delta\tau,\cdots,\Delta\tau_n$，鉴于水面或水下目标的运动比较缓慢，在$T_0$期间$\Delta\tau$不可能有急剧变化，因此可用直线来拟合$\Delta\tau$的变化规律。

根据线性拟合公式(参见§2-5)，得到

$$\hat{\Delta\tau}_k=\overline{\Delta\tau}+\frac{\sum_{i=1}^{n}(\Delta\tau_i-\overline{\Delta\tau})(t_i-\overline{t})}{\sum_{i=1}^{n}(t_i-\overline{t})^2}(t_k-\overline{t}) \tag{9-3-29}$$

其中

$$\left.\begin{aligned}\overline{t}&=\frac{T_0}{2}\\ t_k-\overline{t}&=T_0-\frac{T_0}{2}=\frac{T_0}{2}\\ \overline{\Delta\tau}&=\frac{1}{n}\sum_{i=1}^{n}\Delta\tau_i\end{aligned}\right\} \tag{9-3-30}$$

故得到在k时刻的估计$\hat{\Delta\tau}_k$为

$$\hat{\Delta\tau}_k=\frac{1}{n}\sum_{i=1}^{n}\Delta\tau_i+\frac{\sum_{i=1}^{n}\left(\Delta\tau_i-\frac{1}{n}\sum_{i=1}^{n}\Delta\tau_i\right)\left(t_i-\frac{T_0}{2}\right)}{\sum_{i=1}^{n}\left(t_i-\frac{T_0}{2}\right)^2}\cdot\frac{T_0}{2} \tag{9-3-31}$$

到$k+1$时刻，滑窗T_0移动一步，便得到$k+1$时刻的估计$\hat{\Delta\tau}_{k+1}$，参见图9-3-7。

对于测量数据τ_{12}、τ_{34}进行同样处理，得到$\hat{\tau}_{12k}$和$\hat{\tau}_{34k}$，于是k时刻的θ_{1k}和θ_{2k}的估计如下式所示

$$\hat{\theta}_{1k}=\sin^{-1}\frac{\text{[illegible]}}{l} \tag{9-3-32}$$

$$\hat{\vartheta}_{2k}=\sin^{-1}\frac{c\hat{\tau}_{34k}}{l}\tag{9-3-33}$$

方位估计为

$$\hat{\theta}_k=\frac{\hat{\theta}_{1k}+\hat{\theta}_{2k}}{2}\tag{9-3-34}$$

距离估计为

$$\hat{r}_k=\frac{d\,l\cos^2\hat{\vartheta}_k}{c\,\Delta\hat{\tau}_k}\tag{9-3-35}$$

窗口宽度T_0应适当选取，一方面应对$\Delta\tau$的慢起伏起到抑制作用，另一方面还应保证对运动目标的跟踪。

上述线性拟合方法经过实际试验，取得了一定效果，测距精度有了提高，设计者声称试验样机的测距误差约在10%左右。

事实上，不采用对τ_D的直接平滑处理，而是对距离观测值采用下章中的自适应算法，可以得到更好的结果。

§9-4 在鱼雷航迹测量中的应用

为了检验鱼雷是否按照指定的轨道航行，需要实时跟踪鱼雷的航迹，采用卡尔曼滤波器可以给出较精确的航迹估计，对鱼雷的定位测量装置如图9-4-1所示。

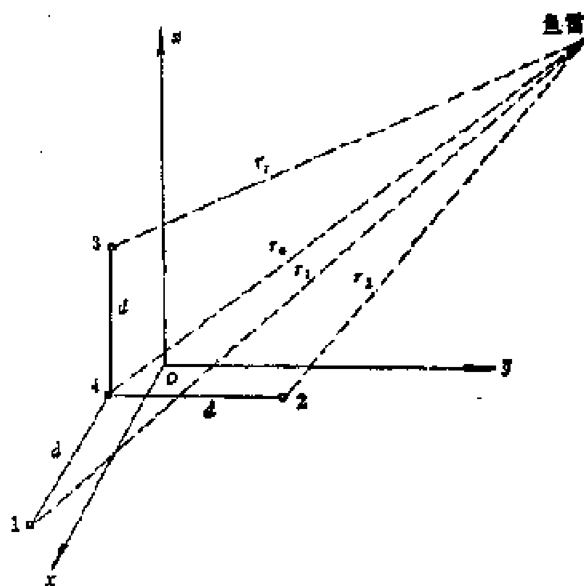


图9-4-1 鱼雷航迹测量装置

用四个水听器接装在一个立方体的四个顶点上，鱼雷上接装有小型发射机，每隔T_0秒发出一个声脉冲。鱼雷上及水听器接收机上装有同步时钟，在下水前进行对准，同步时钟有足够的稳定性，保证在试验时间内不会失去同步，因而鱼雷发射声脉冲的时间是已知的。每个水听器测到的时延与鱼雷至该水听器的距离相对应，取立方体的中心为坐标系的原点，于是各水听器至鱼雷的距离为

$$\left.\begin{aligned}
r_0&=\left[\left(x+\frac{d}{2}\right)^2+\left(y+\frac{d}{2}\right)^2+\left(z+\frac{d}{2}\right)^2\right]^{\frac{1}{2}}\\
r_1&=\left[\left(x-\frac{d}{2}\right)^2+\left(y+\frac{d}{2}\right)^2+\left(z+\frac{d}{2}\right)^2\right]^{\frac{1}{2}}\\
r_2&=\left[\left(x+\frac{d}{2}\right)^2+\left(y-\frac{d}{2}\right)^2+\left(z+\frac{d}{2}\right)^2\right]^{\frac{1}{2}}\\
r_3&=\left[\left(x+\frac{d}{2}\right)^2+\left(y-\frac{d}{2}\right)^2+\left(z-\frac{d}{2}\right)^2\right]^{\frac{1}{2}}
\end{aligned}\right\}\tag{9-4-1}$$

再来考察鱼雷的运动方式。在制导鱼雷上通常装有寻的声纳，鱼雷入水后首先作圆周运动进行搜索，当发现目标后，即对准目标行进，由于惰性作用，将按蛇行方式行进。鱼雷的深度，则由鱼雷上按装的定深装置设定，鱼雷的运动方式如图9-4-2所示。

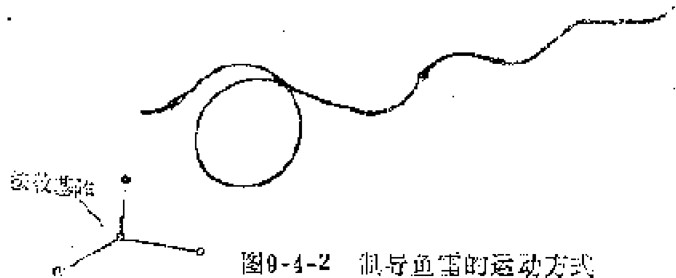


图9-4-2　制导鱼雷的运动方式

根据以上讨论，就可以建立适当的数学模型，并构造相应的卡尔曼滤波器。

选定状态矢量为

$$X=[x,\ \dot{x},\ y,\ \dot{y},\ z]^T \tag{9-4-2}$$

因为鱼雷在航行过程中，基本上按某个设定的深度航行，而且我们对深度的变化率不发生兴趣，所以它没有包含在状态矢量中。

列出状态方程如下：

$$\begin{pmatrix} x \\ \dot{x} \\ y \\ \dot{y} \\ z \end{pmatrix}_{k+1}=\begin{pmatrix} 1 & T & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & T & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ \dot{x} \\ y \\ \dot{y} \\ z \end{pmatrix}_k+\begin{pmatrix} \frac{T^2}{2} & 0 & 0 \\ T & 0 & 0 \\ 0 & \frac{T^2}{2} & 0 \\ 0 & T & 0 \\ 0 & 0 & T \end{pmatrix}\begin{pmatrix} \ddot{w}_x \\ \ddot{w}_y \\ \dot{w}_z \end{pmatrix}_k \tag{9-4-3}$$

其中T为采样周期，$\ddot{w}_x$和$\ddot{w}_y$为两个加速度扰动噪声分量，$\dot{w}_z$为速度扰动噪声分量，假设它们是零均值白噪声。

再列出测量方程如下：

$$\left.\begin{aligned} cT_0&=\left[\left(x+\frac{d}{2}\right)^2+\left(y+\frac{d}{2}\right)^2+\left(z+\frac{d}{2}\right)^2\right]^{\frac{1}{2}}+V_0 \\ cT_1&=\left[\left(x-\frac{d}{2}\right)^2+\left(y+\frac{d}{2}\right)^2+\left(z+\frac{d}{2}\right)^2\right]^{\frac{1}{2}}+V_1 \\ cT_2&=\left[\left(x+\frac{d}{2}\right)^2+\left(y-\frac{d}{2}\right)^2+\left(z+\frac{d}{2}\right)^2\right]^{\frac{1}{2}}+V_2 \\ cT_3&=\left[\left(x+\frac{d}{2}\right)^2+\left(y+\frac{d}{2}\right)^2+\left(z-\frac{d}{2}\right)^2\right]^{\frac{1}{2}}+V_3 \end{aligned}\right\} \tag{9-4-4}$$

其中V_0、V_1、V_2、V_3为接收阵各元的测距误差，记

$$T_k=\begin{pmatrix} cT_0 \\ cT_1 \\ cT_2 \\ cT_3 \end{pmatrix}_k \qquad V_k=\begin{pmatrix} V_0 \\ V_1 \\ V_2 \\ V_3 \end{pmatrix}_k$$

则测量方程可写成

$$T_k=M(X_k)+V_k \tag{9-4-5}$$

再来估算噪声统计性质。对于各基元的测量噪声，可以合理地假定它们是零均值白噪

声，互相独立，方差相等，因此有

$$R_k = E[v_k v_k^T] = \sigma_k^2 I_6 \tag{9-4-6}$$

由于水声信道的一些特点，如声速起伏，介质不均匀等因素引起的测距误差大约与距离成比例，一般可取$\sigma_k = 0.05r_k \sim 0.10r_k$，因为$r_k$是待估计此，可以用$\hat{r}_{k/k-1}$代替之。

$$\hat{r}_{k/k-1} = [\hat{x}^2_{k/k-1} + \hat{y}^2_{k/k-1} + \hat{z}^2_{k/k-1}]^{1/2}$$

至于状态噪声矢量$W_k = [\ddot{w}_{xk}, \ddot{w}_{yk}, \dot{w}_{hk}]^T$的统计性质可从鱼雷的操纵舵的控制力矩的概率分布确定之。

按照以上讨论得出的数学模型，我们就可设计相应的卡尔曼滤波器，滤波器的设计可采用不同的方案。

一、广义卡尔曼滤波器方案

重写信号模型如下：

$$X_{k+1} = \Phi_{k+1/k} X_k + \Gamma W_k \tag{9-4-7}$$

$$T_{k+1} = M(X_{k+1}) + V_{k+1} \tag{9-4-8}$$

按照广义卡尔曼滤波器设计方法，首先求出线性化测量矩阵H_{k+1}

$$H_{k+1} = \frac{\partial M(X)}{\partial X}\bigg|_{X=\hat{X}_{k+1/k}}$$

$$= \begin{pmatrix} \frac{x+\frac{d}{2}}{r_0} & 0 & \frac{y+\frac{d}{2}}{r_0} & 0 & \frac{z+\frac{d}{2}}{r_0} \\ \frac{x-\frac{d}{2}}{r_1} & 0 & \frac{y+\frac{d}{2}}{r_1} & 0 & \frac{z+\frac{d}{2}}{r_1} \\ \frac{x+\frac{d}{2}}{r_2} & 0 & \frac{y-\frac{d}{2}}{r_2} & 0 & \frac{z+\frac{d}{2}}{r_2} \\ \frac{x+\frac{d}{2}}{r_3} & 0 & \frac{y+\frac{d}{2}}{r_3} & 0 & \frac{z-\frac{d}{2}}{r_3} \end{pmatrix}_{X=\hat{X}_{k+1/k}} \tag{9-4-9}$$

于是可以设计出广义卡尔曼滤波器如下：

$$\hat{X}_{k+1/k} = \Phi \hat{X}_k \tag{9-4-10}$$

$$P_{k+1/k} = \Phi P_k \Phi^T + Q_k^* \tag{9-4-11}$$

$$K_{k+1} = P_{k+1/k} H_{k+1}^T [H_{k+1} P_{k+1/k} H_{k+1}^T + R_{k+1}^*]^{-1} \tag{9-4-12}$$

$$\hat{X}_{k+1} = \hat{X}_{k+1/k} + K_{k+1}[T_{k+1} - M(\hat{X}_{k+1/k})] \tag{9-4-13}$$

$$P_{k+1} = (I - K_{k+1}H_{k+1})P_{k+1/k}(I - K_{k+1}H_{k+1})^T + K_{k+1}R_{k+1}^* K_{k+1}^T \tag{9-4-14}$$

其中Q_k^*、R_k^*为原有噪声方差Q_k、R_k再加上线性化误差引入的成分。

二、顺序卡尔曼滤波器方案

四个测量值是顺序地送入滤波器的，以T_i表示之，$i = 1, 2, 3, 4$。

假定已得出k时刻的滤波值$\hat{X}_k$和P_k，以及一步预报$\hat{X}_{k+1/k}$和$P_{k+1/k}$，我们并不等四个

测量值全部进入后再进行下一轮计算，而是每输入一个数据就计算一次。当第一个 cT_1 送入时，就计算 K_1、$\hat{X}_1$、P_1，并以 $\hat{X}_1$、P_1 为预报值；第二个数据 cT_2 输入时，就计算 K_2、$\hat{X}_2$、P_2。依此类推，当第四个数据 cT_4 输入时，计算出 K_4、$\hat{X}_4$、P_4，把 $\hat{X}_4$、P_4 作为 $k+1$ 时刻状态估计值，令

$$\hat{X}_{k+1} = \hat{X}_4, \qquad P_{k+1} = P_4$$

然后进行下一轮计算，这种顺序地利用测量值的方法，称为顺序卡尔曼滤波方法。

计算 K_i、$\hat{X}_i$、P_i 的方法如下：由卡尔曼滤波器公式可以验证

$$K_{icol} = \frac{P_{k+1/k} H_{irow}^T}{H_{irow} P_{k+1/k} H_{irow}^T + R_{ii}} \tag{9-4-15}$$

其中 H_{irow} 为 H 矩阵中第 i 横行；

K_{icol} 为 K 矩阵中第 i 纵列；

R_{ii} 为 R 矩阵中第 ii 元素。

令 $c\tilde{T}_i$ 代表本次测量值与预报测量值之差

$$c\tilde{T}_i = cT_i - M_i(\hat{X}_{i/i-1}) \tag{9-4-16}$$

于是，得第 i 次测量的更新为

$$\hat{X}_i = \hat{X}_{i-1} + K_{icol}\, c\tilde{T}_i \tag{9-4-17}$$

$$P_i = (1 - K_{icol} H_{irow}) P_{i-1} \tag{9-4-18}$$

而当 $i = 1$ 时，应取

$$\hat{X}_1 = \hat{X}_{k+1/k} + K_{1col}\, c\tilde{T}_1 \tag{9-4-19}$$

$$P_1 = (1 - K_{1col} H_{1row}) P_{k+1/k} \tag{9-4-20}$$

其中 $\hat{X}_{k+1/k} = \Phi \hat{X}_k$

$P_{k+1/k} = \Phi P_k \Phi^T + \Gamma Q_k \Gamma^T$

到下一个采样周期，重复上述计算步骤。顺序卡尔曼滤波方法的优点在于不需要计算逆矩阵，这对于提高计算速度和数值计算精度都是有益的。其次，当四个测量值中有某个漏报，可以跳过这一步，对于滤波结果不会有大的影响。

三、线性卡尔曼滤波方案

因为各水听器至鱼雷的距离为

$$\left.\begin{aligned}
r_0 &= \left[\left(x + \frac{d}{2}\right)^2 + \left(y + \frac{d}{2}\right)^2 + \left(z + \frac{d}{2}\right)^2\right]^{\frac{1}{2}} \\
r_1 &= \left[\left(x - \frac{d}{2}\right)^2 + \left(y + \frac{d}{2}\right)^2 + \left(z + \frac{d}{2}\right)^2\right]^{\frac{1}{2}} \\
r_2 &= \left[\left(x + \frac{d}{2}\right)^2 + \left(y - \frac{d}{2}\right)^2 + \left(z + \frac{d}{2}\right)^2\right]^{\frac{1}{2}} \\
r_3 &= \left[\left(x + \frac{d}{2}\right)^2 + \left(y + \frac{d}{2}\right)^2 + \left(z - \frac{d}{2}\right)^2\right]^{\frac{1}{2}}
\end{aligned}\right\} \tag{9-4-21}$$

由上述方程组解出

$$
\left.\begin{aligned}
x &= \frac{r_0^2 - r_1^2}{2d} = \frac{c^2}{2d}\,(t_0^2 - t_1^2) \\
y &= \frac{r_0^2 - r_2^2}{2d} = \frac{c^2}{2d}(t_0^2 - t_2^2) \\
z &= \frac{r_0^2 - r_3^2}{2d} = \frac{c^2}{2d}(t_0^2 - t_3^2)
\end{aligned}\right\} \qquad (9\text{-}4\text{-}22)
$$

令

$$t_0 - t_1 = \Delta t_1$$

则有

$$
\begin{aligned}
x &= \frac{c^2}{2d}(t_0 + t_1)(t_0 - t_1) \\
&= \frac{c^2}{2d}(2t_0 - \Delta t_1)\Delta t_1 \\
&= \frac{c^2}{2d}2t_0\Delta t_1 - \frac{c^2}{2d}\Delta t_1^2
\end{aligned} \qquad (9\text{-}4\text{-}33)
$$

同理，令$\Delta t_2 = t_0 - t_2$，$\Delta t_3 = t_0 - t_3$，可解出

$$y = \frac{c^2}{2d}2t_0\Delta t_2 - \frac{c^2}{2d}\Delta t_2^2 \qquad (9\text{-}4\text{-}24)$$

$$z = \frac{c^2}{2d}2t_0\Delta t_3 - \frac{c^2}{2d}\Delta t_3^2 \qquad (9\text{-}4\text{-}25)$$

因为$\Delta t_i \ll 2t_0$，略去高阶项，得

$$
\left.\begin{aligned}
x &\approx \frac{c^2}{d}t_0\Delta t_1 \\
y &\approx \frac{c^2}{d}t_0\Delta t_2 \\
z &\approx \frac{c^2}{d}t_0\Delta t_3
\end{aligned}\right\} \qquad (9\text{-}4\text{-}26)
$$

由k时刻的测量值t_{0k}、Δt_{1k}、Δt_{2k}、Δt_{3k}可算出目标坐标的测量值，并用符号ρ_{xk}、ρ_{yk}、ρ_{zk}表示之。考虑到t_{0k}、Δt_{1k}、Δt_{2k}、Δt_{3k}中均包含有测量误差，于是可列出测量方程如下：

$$\rho_{xk} = \frac{c^2}{d}t_{0k}\Delta t_{1k} + v_{xk}$$

$$\rho_{yk} = \frac{c^2}{d}t_{0k}\Delta t_{2k} + v_{yk}$$

$$\rho_{zk} = \frac{c^2}{d}t_{0k}\Delta t_{3k} + v_{zk}$$

写成矢量形式，得

$$
\begin{pmatrix} \rho_x \\ \rho_y \\ \rho_z \end{pmatrix}_k = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \\ y \\ \dot{y} \\ z \end{pmatrix}_k + \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}_k \qquad (9\text{-}4\text{-}27)
$$

系统的状态方程则为

$$\begin{pmatrix} x \\ \dot{x} \\ y \\ \dot{y} \\ z \end{pmatrix}_{k+1} = \begin{pmatrix} 1 & T & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & T & 0 \\ 0 & 0 & 0 & 1 & T \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \\ y \\ \dot{y} \\ z \end{pmatrix}_k + \begin{pmatrix} \frac{T^2}{2} & 0 & 0 \\ T & 0 & 0 \\ 0 & \frac{T^2}{2} & 0 \\ 0 & T & 0 \\ 0 & 0 & T \end{pmatrix} \begin{pmatrix} \ddot{w}_x \\ \ddot{w}_y \\ \dot{w}_z \end{pmatrix} \quad (9\text{-}4\text{-}28)$$

其中$\ddot{w}_x$和$\ddot{w}_y$分别是x和y方向上的加速度噪声，$\dot{w}_z$是z方向上的速度噪声。

由于本方案中的模型是线性的，故可以采用线性卡尔曼滤波算法。

§9-5 在最优控制中的应用

关于最优控制是一个比较复杂的问题，超出了本书范围，这里主要讨论一下卡尔曼滤波器在最优控制中有什么应用。

先考察一个确定性过程，其模型为

$$X_{k+1} = \Phi_{k+1/k} X_k + B_k U_k \quad (9\text{-}5\text{-}1)$$

$$Z_{k+1} = H_{k+1} X_{k+1} \quad (9\text{-}5\text{-}2)$$

要求在时间区间$(0,\ NT)$内加上控制序列

$$U_i = U_0,\ U_1,\ \cdots,U_{N-1}$$

使二次型性能指标

$$J_N = \sum_{i=1}^{N} [X_i^T A_i X_i + U_{i-1}^T D_{i-1} U_{i-1}] \quad (9\text{-}5\text{-}3)$$

达到最小，其中A_i、D_i为根据某种要求设定的矩阵，如果低能量消耗、最短时间等。设A_i为$n\times n$维矩阵，D_i为$m\times m$维矩阵，X为n维矩阵，U为m维矢量。

根据最优控制理论得知，当状态X_k为完全可观测时，最优控制U_k^*可表达为

$$U_k^* = L_k X_k \quad (9\text{-}5\text{-}4)$$

其中L_k称为最优反馈增益，可由动态规划理论导出的递推算法给出。

尽管这一问题已在理论上得到解决，但实际上，过程的状态变化并非如此确定，对状态的测量也受到噪声干扰，因而，过程的模型应由下式来表达

$$X_{k+1} = \Phi_{k+1/k} X_k + B_k U_k + \Gamma_k W_k \quad (9\text{-}5\text{-}5)$$

$$Z_{k+1} = H_{k+1} X_{k+1} + V_{k+1} \quad (9\text{-}5\text{-}6)$$

对于随机过程的最优控制问题称为随机最优控制。由于性能指标也是随机量，应当用其均值表示，即有

$$E[J_N] = E\left\{\sum_{i=1}^{N} [X_i^T A_i X_i + U_{i-1}^T D_{i-1} U_{i-1}]\right\} \quad (9\text{-}5\text{-}7)$$

现在的问题是选取什么样的控制序列U_i才能使$E[J_N]$达到最小。要解决这一问题，我们自然会想到，利用该时刻之前的测量序列，对状态矢量作出估计$\hat{X}_k$，然后令最优控制序列为

$$U_k^* = L_k \hat{X}_k \tag{9-5-8}$$

U_k^*是否为上述随机过程的最优控制序列，是需要加以证明的。如果回答是肯定的，那么，随机最优控制就可分为状态估计与最优控制两个步骤来实现，也就是说，不改变原来确定性最优控制系统的结构，只要加上一个状态估计器，即卡尔曼滤波器，就可实现随机过程的最优控制，如图 9-5-1 所示。

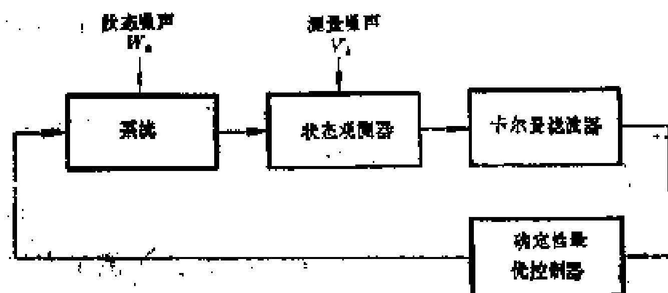


图9-5-1　随机过程最优控制结构

把随机最优控制问题分成两步求解称为分离原理。对于线性二次型问题，当状态噪声 W_k 和测量噪声V_k均为白噪声时，已经证明分离原理成立，这是一类相当广泛的工程问题的数学模型，而卡尔曼滤波器作为状态估计器在最优控制中得到广泛应用。

§9-6　在AR谱估计中的应用

平稳随机时间序列可以看成是由平稳白噪声系列驱动一个线性时不变滤波器的输出。平稳随机时间序列的p阶AR模型可表达为

$$X_k = a_1 X_{k-1} + a_2 X_{k-2} + \cdots + a_p X_{k-p} + Gv_k$$

$$= \sum_{i=1}^{p} a_i X_{k-i} + Gv_k \tag{9-6-1}$$

其中G为常数，v_k为平稳零均值白噪声，方差等于1。信号的AR模型结构如图9-6-1所示。

当滤波器参数a_i已知时，容易求得此平稳随机系列的功率谱估计。

对(9-6-1)式取变换，可得出AR滤波器的传递函数

$$H(z) = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} \tag{9-6-2}$$

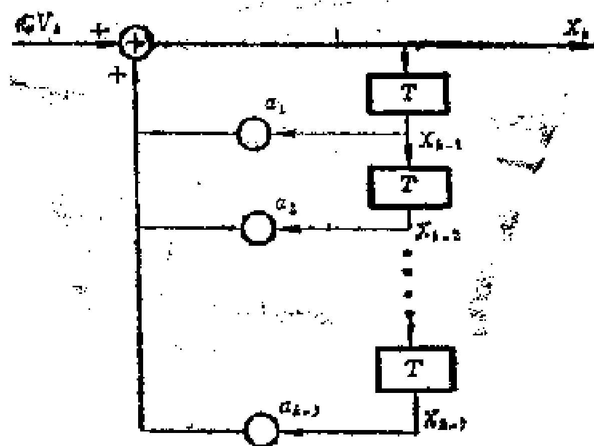


图9-6-1　平稳随机序列的模型

输出功率谱$S_x(z)$与输入功率谱$S_v(z)$之间的关系为

$$S_x(z) = H(z) H(z^{-1}) S_v(z) \tag{9-6-3}$$

已知$S_v(z) = G^2$，令$z = e^{-j\omega T}$，T为采样周期，得输出功率谱为

$$S_x(\omega) = \frac{G^2}{\left|1 - \sum_{i=1}^{p} a_i \exp(-j\omega T i)\right|^2} \tag{9-6-4}$$

由(9-6-4)式可见，噪声方差G^2不影响$S_x(\omega)$的分布形状而只影响幅度。

卡尔曼滤波法求参数a_i的估计

通常情况下参数a_i为未知，需要通过观测值来出a_i的估计。的估计方法有好几种，本章中提出用卡尔曼滤波方法，其它方法将在第十二章中讨论。

定义测量矩阵H_k

$$H_k = [X_{k-1},\ X_{k-2},\ \cdots,\ X_{k-p}] \tag{9-6-5}$$

再定义参数矢量 θ

$$\theta = [a_1,\ a_2,\ \cdots,\ a_p]^T \tag{9-6-6}$$

则(9-6-1)式可写成

$$X_k = H_k\theta_k + Gv_k \tag{9-6-7}$$

因为滤波器为时不变的，故θ_k为常值矢量，有

$$\theta_k = \theta_{k-1} \tag{9-6-8}$$

(9-6-7)、(9-6-8) 式构成了参数矢量θ的状态变量模型，其中 (9-6-7) 式为测量方程，(9-6-8) 式为状态方程。于是我们可以应用卡尔曼滤波求θ的估计，相应的卡尔曼滤波公式如下：

$$K_{k+1} = P_k H_{k+1}^T [H_{k+1} P_k H_{k+1}^T + G^2]^{-1} \tag{9-6-9}$$

$$\hat{\theta}_{k+1} = \hat{\theta}_k + K_{k+1}[X_{k+1} - H_{k+1}\hat{\theta}_k] \tag{9-6-10}$$

$$P_{k+1} = [1 - K_{k+1}H_{k+1}]P_k \tag{9-6-11}$$

给定初值$\hat{\theta}_0$和P_0，即可启动滤波器，滤波器的收敛速度与G值大小有关，G小时收敛快，G大时收敛慢。

需要指出，本模型是无状态噪声的，当信号略有非平稳时，滤波器容易发散，为了防止发散，可在状态方程中加入伪噪声，令

$$\theta_{k+1} = \theta_k + \xi_k \tag{9-6-12}$$

其中ξ_k为平稳零均值白噪声，令其方差等于Q，则相应的卡尔曼滤波器为

$$\hat{\theta}_{k+1/k} = \theta_k \tag{9-6-13}$$

$$P_{k+1/k} = P_k + Q \tag{9-6-14}$$

$$K_{k+1} = P_{k+1/k} H_{k+1}^T [H_{k+1} P_{k+1/k} H_{k+1}^T + G^2]^{-1} \tag{9-6-15}$$

$$\hat{\theta}_{k+1} = \hat{\theta}_{k+1/k} + K_{k+1}[X_{k+1} - H_{k+1}\hat{\theta}_{k+1/k}] \tag{9-6-16}$$

$$P_{k+1} = [1 - K_{k+1}H_{k+1}]P_{k+1/k} \tag{9-6-17}$$

Q_k不应取得太大，以免影响估计精度。经过多次迭代计算后，就可得到参数矢量θ的估计

$$\hat{\theta} = [\hat{a}_1,\ \hat{a}_2, \cdots, \hat{a}_p] \tag{9-6-18}$$

把$\hat{a}_i$代入(9-6-4)式，就得到随机序列X_k的功率谱估计

$$\hat{S}_x(\omega) = \frac{1}{\left|1 - \sum_{i=1}^{p} \hat{a}_i \exp(-j\omega T i)\right|^2} \tag{9-6-19}$$

还可以得到噪声功率的估计

$$\hat{G}^2 = \left(X_k - \sum_{i=1}^{p} \hat{a}_i X_{k-i}\right)^2 \tag{9-6-20}$$

必须指出，求平稳随机序列的功率谱估计尚存在两个问题没有解决，需要进一步研究。一是测量矩阵H_k是由测量值组成的，而测量是有噪声的，这一点不同于常规的卡尔曼滤波器。在那里测量矩阵是确定量，实际上是修改了卡尔曼滤波理论，修改后的影响如何是需要研究的。二是先验知识缺乏的情况下，如何决定AR模型的阶数p，下面分别讨论之。

参数估计$\hat{a}_i$的性能

根据第七章的讨论，我们知道卡尔曼滤波与最优加权最小二乘估计有等价关系，因而可以从最优加权最小二乘估计出发来研究卡尔曼滤波参数估计的性能，即研究参数估计$\hat{a}_i$的偏量与方差。

设有平稳AR(p)模型

$$X_k = a_1 X_{k-1} + a_2 X_{k-2} + \cdots + a_p X_{k-p} + Gv_k$$

从$k=p+1$次开始，列出各次测量模型，有

$$X_{p+1} = a_1 X_p + a_2 X_{p-1} + \cdots + a_p X_1 + Gv_{p+1}$$

$$X_{p+2} = a_1 X_{p+1} + a_2 X_p + \cdots + a_p X_2 + Gv_{p+2}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$X_{k-1} = a_1 X_{k-2} + a_2 X_{k-3} + \cdots + a_p X_{k-p-1} + Gv_{k-1}$$

$$X_k = a_1 X_{k-1} + a_2 X_{k-2} + \cdots + a_p X_{k-p} + Gv_k$$

写成矩阵形式，得

$$\begin{pmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_{k-1} \\ X_k \end{pmatrix} = \begin{pmatrix} X_p & X_{p-1} & \cdots & X_1 \\ X_{p+1} & X_p & \cdots & X_2 \\ \cdots & \cdots & \cdots & \cdots \\ X_{k-2} & X_{k-3} & \cdots & X_{k-p-1} \\ X_{k-1} & X_{k-2} & \cdots & X_{k-p} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{p-1} \\ a_p \end{pmatrix} + G \begin{pmatrix} v_{p+1} \\ v_{p+2} \\ \vdots \\ v_{k-1} \\ v_k \end{pmatrix} \tag{9-6-21}$$

记

$$\mathbf{X}_k = \begin{pmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_{k-1} \\ X_k \end{pmatrix}, \quad \boldsymbol{\Theta}_k = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{p-1} \\ a_p \end{pmatrix}, \quad \boldsymbol{V}_k = \begin{pmatrix} v_{p+1} \\ v_{p+2} \\ \vdots \\ v_{k-1} \\ v_k \end{pmatrix}$$

$$\mathbf{H}_k = \begin{pmatrix} H_{p+1} \\ H_{p+2} \\ \vdots \\ H_{k-1} \\ H_k \end{pmatrix} = \begin{pmatrix} X_p & X_{p-1} & \cdots X_1 \\ X_{p+1} & X_p & \cdots X_2 \\ \cdots & \cdots & \cdots \\ X_{k-2} & X_{k-3} & \cdots X_{k-p-1} \\ X_{k-1} & X_{k-2} & \cdots X_{k-p} \end{pmatrix}$$

于是有矢量方程

$$\mathbf{X}_k = \mathbf{H}_k \boldsymbol{\Theta}_k + G\boldsymbol{V}_k \tag{9-6-22}$$

对Θ_k的最优加权最小二乘估计为

$$\hat{\Theta}_k = (\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T \mathbf{X}_k \tag{9-6-23}$$

$$P_k = G^2 (\mathbf{H}_k^T \mathbf{H}_k)^{-1} \tag{9-6-24}$$

最优加权最小二乘估计的递推形式为

$$\hat{\theta}_{k+1}=\hat{\theta}_k+K_{k+1}[X_{k+1}-H_{k+1}\hat{\theta}_k] \tag{9-6-25}$$

$$K_{k+1}=P_kH_{k+1}^T[H_{k+1}P_kH_{k+1}^T+G^2]^{-1} \tag{9-6-26}$$

$$P_{k+1}=[I-K_{k+1}H_{k+1}]P_k \tag{9-6-27}$$

见于卡尔曼滤波与最优加权最小二乘估计的等价性，(9-6-9)～(9-6-11)式与(9-6-23)、(9-9-24)式或(9-6-25)～(9-6-27)式等价，我们从(9-6-23)、(9-6-24)式出发来考察参数$\hat{\theta}_k$的性能。

把(9-6-22)式代入(9-6-23)式中，得

$$\begin{aligned}\hat{\theta}_k&=(\mathrm{H}_k^T\mathrm{H}_k)^{-1}\mathrm{H}_k^T(\mathrm{H}_k\theta_k+GV_k)\\&=\theta_k+G(\mathrm{H}_k^T\mathrm{H}_k)^{-1}\mathrm{H}_k^TV_k\end{aligned} \tag{9-6-28}$$

把H_k、V_k用其分量表示，有

$$\begin{aligned}\hat{\theta}_k&=\theta_k+\left(\sum_{i=p+1}^{k}H_i^TH_i\right)^{-1}\left(G\sum_{i=p+1}^{k}H_i^Tv_i\right)\\&=\theta_k+\left(\frac{1}{k-1}\sum_{i=p+1}^{k}H_i^TH_i\right)^{-1}\left(\frac{G}{k-p}\sum_{i=p+1}^{k}H_i^Tv_i\right)\end{aligned} \tag{9-6-29}$$

已知v_i为零均值白噪声，故v_i只与t_i之后的信号相关，而与t_i之前的信号不相关，故有

$$E[H_i^Tv_i]=E[X_{i-1},\ X_{i-2},\cdots,X_{i-p})^Tv_i]=0 \tag{9-6-30}$$

又已知X_k为零均值平稳序列，故相关函数等于协方差函数，有

$$E[X_{i-j}X_{i-s}]=r_{s-j} \tag{9-6-31}$$

$$E[H_i^TH_i]=\begin{pmatrix}r_0 & r_1 & \cdots & r_{p-1}\\ r_1 & r_0 & \cdots & r_{p-2}\\ \cdots & \cdots & \cdots & \cdots\\ r_{p-1} & r_{p-2} & \cdots & r_p\end{pmatrix}=R \tag{9-6-32}$$

R为对称正定矩阵，其逆阵存在。在上述条件下，推知在以概率1和均方收敛意义下，有

$$当k\to\infty时，\ \frac{1}{k-p}\sum_{i=p+1}^{k}H_i^Tv_i\to E[H_i^Tv_i]=0 \tag{9-6-33}$$

$$当k\to\infty时，\ \frac{1}{k-p}\sum_{i=p+1}^{k}H_i^TH_i\to E[H_i^TH_i]=R \tag{9-6-34}$$

所以，由(9-6-29)式可推知，$\hat{\theta}_k$以概率1和均方收敛到θ_k。

由(9-6-24)式，得

$$\begin{aligned}P_k&=G^2[\mathrm{H}_k^T\mathrm{H}_k]^{-1}=G^2\left[\sum_{i=p+1}^{k}H_i^TH_i\right]^{-1}\\&=\frac{G^2}{k-p}\left[\frac{1}{k-p}\sum_{i=p+1}^{k}H_i^TH_i\right]^{-1}\end{aligned} \tag{9-6-35}$$

根据(9-6-34)式，可推知在以概率1和均方收敛意义下，有

$$当k\to\infty时，\ P_k\to\frac{G^2}{k-p}R^{-1}=0 \tag{9-6-36}$$

考察(9-6-33)、(9-6-34)式，实质上是在求相关函数的估计。对于有限数据长度，相关函数的估计有两种求法。

设有实平稳随机序列，数据长度是N。其自相关序列为

$$\Phi_x(m)=E[X_n X_{n+m}],\quad |m|<N \tag{9-6-37}$$

自相关序列的估计之一是取

$$\check{\Phi}_x(m)=\frac{1}{N-m}\sum_{n=0}^{N-|m|-1} X_n X_{n+m} \tag{9-6-38}$$

不难证明，$\check{\Phi}_x(m)$是$\Phi_x(m)$的无偏估计。

自相关序列的估计之二是取

$$\hat{\Phi}_x(m)=\frac{1}{N}\sum_{n=0}^{N-|m|-1} X_n X_{n+m} \tag{9-6-39}$$

不难证明

$$E[\hat{\Phi}_x(m)]=\frac{N-|m|}{N}\Phi_x(m) \tag{9-6-40}$$

因此，$\hat{\Phi}_x(m)$是$\Phi_x(m)$的有偏估计，但又是渐近无偏的。

显然(9-6-33)、(9-6-34)式的估计都是有偏估计，但又是渐近无偏的。

再考察(9-6-35)式，[　]内因子是R的有偏估计，也又是渐近无偏的，故P_k渐近地收敛于零(未考虑伪噪声)。

常规的卡尔曼滤波器中得到的是线性无偏最小方差估计，在这里由于修改了卡尔曼滤波理论，得到的估计不再是无偏和最小方差的。但是，只要迭代步数足够长时仍然能得到好的参数估计。

卡尔曼滤波法参数估计是递推算法。当信号是非平稳时，即参数矢量Θ_k是时变的，我们可以把变化量看成是状态噪声，只要计算速度跟得上，就能很好跟踪Θ_k的变化。所以，卡尔曼滤波方法对时变谱的跟踪将比其他谱估计方法(参见第十二章)容易。

AR(p)模型阶数p的确定

确定未知模型阶数也有好几种方法，参见第十二章，本节中只讲一种F检验判定法。

设平稳序列X_k用p阶AR模型表示

$$X_k=\hat{a}_1X_{k-1}+\hat{a}_2X_{k-2}+\cdots+\hat{a}_pX_{k-p}+\text{残差} \tag{9-6-41}$$

残差的平方和为

$$RSS=\sum_{k=p}^{N}(X_k-\hat{a}_1X_{k-1}-\cdots-\hat{a}_pX_{k-p})^2 \tag{9-6-42}$$

N为信号采样数目，$N>p$。

记　$A_0=p$阶模型AR(p)的残差平方和，

$A_1=p-1$阶模型AR($p-1$)的残差平方和。

建立如下判决规则：

H_0：AR($p-1$)模型是合适的　$a_p=0$，

H_1：AR($p-1$)模型是不合适的$a_p\neq0$。

则统计量

$$F=\frac{A_1-A_0}{A_0}(N-p) \tag{9-6-43}$$

服从$F(1, N-p)$分布。给定显著性水平b，例如取$b=0.02$，查F分布表，可得临界值F_b。

若$F \geqslant F_b$，则拒绝H_0，即AR($p-1$)不合适。

若$F < F_b$，则接受H_0，即AR($p-1$)合适。

对已给定的观测序列，从低阶开始，逐步增大p值，对相邻两个AR模型用上述F检验法进行判决，直到找到合适的阶数p为止。

第十章　自适应卡尔曼滤波

从第七章得知，卡尔曼滤波为线性无偏最小方差估计，但要求信号模型正确，噪声统计量已知，实际上往往不具备。自七十年以来，自适应卡尔曼滤波算法的研究受到十分重视，并且还处在继续发展之中。本章研究自适应卡尔曼滤波的一些主要成果。

设计不正确的卡尔曼滤波器可能是发散的，造成发散的主要原因在于所用模型不正确，因此原因造成的滤波发散称为模型发散。造成滤波发散的另一原因是数值计算中的截尾，称为数值发散。本章还研究这两种滤波发散的机理及其抑制方法。

§10-1　模型发散及其抑制方法

迄今为止，我们一直假定所采用的数学模型是正确的，状态方程与测量方程描述的是同一时间过程，并由此导出了最佳滤波递推算法——卡尔曼滤波器。但在实际上，状态方程是我们根据对状态变化规律的先验知识作出的数学模型。由于先验知识的不足，或者是为了简化计算而有意简略，使状态方程常常不符合实际情况。而测量虽然有噪声，但尚能粗略地反映实际情况，所以这时状态方程与测量方程描述的不是同一个状态变化规律，这种不一致性轻微时，导致估计误差增大，严重时估计结果完全偏离真实状态，这种现象称为模型发散。意思是说，由于模型不准确而造成了滤波器发散。下面就来研究模型发散的机理及其克服方法。

模型发散机理

我们举一个例子来看滤波器的模型发散现象。

例10-1-1　有一个气球以恒速 a 向上升起，这时它的状态方程和测量方程可以表达为

$$x_{k+1} = x_k + aT \tag{10-1-1}$$

$$z_{k+1} = x_{k+1} + v_{k+1} = x_k + aT + v_{k+1} \tag{10-1-2}$$

其中x_k为k时刻气球上升高度，v_k为测量噪声，并设v_k的均值为零，方差为σ^2。

如果我们误认为气球是静止的，静止气球的状态方程为

$$x_{k+1} = x_k \tag{10-1-3}$$

可见，这个状态方程描述的不是气球高度，因而是错误的。在错误模型下的状态估计如下。

解：由题意可知$\Phi_{k+1,k} = 1$，$H_k = 1$，$Q_k = 0$，$R_k = \sigma^2$ (10-1-4)

给定初值为$\hat{x}_0$和P_0。为方便计，我们采用逆协方差形式的卡尔曼滤波器，滤波器算法如下：

滤波　$$\hat{x}_{k+1} = \hat{x}_k + K_{k+1}[z_{k+1} - \hat{x}_k] \tag{10-1-5}$$

增益　$$K_{k+1} = P_{k+1}/\sigma^2 \tag{10-1-6}$$

方差　$$P_{k+1}^{-1} = P_k^{-1} + 1/\sigma^2 \tag{10-1-7}$$

代入初始值，启动滤波器，得

$$P_1^{-1}=\frac{1}{P_0}+\frac{1}{\sigma^2}$$

$$P_2^{-1}=P_1^{-1}+\frac{1}{\sigma^2}=\frac{1}{P_0}+\frac{2}{\sigma^2}$$

………………

$$P_k^{-1}=\frac{1}{P_0}+\frac{k}{\sigma^2}$$

或
$$P_k=\frac{P_0\sigma^2}{\sigma^2+kP_0} \tag{10-1-8}$$

把此结果代入(10-1-6)式，得增益

$$K_k=\frac{P_k}{\sigma^2}=\frac{P_0}{\sigma^2+kP_0} \tag{10-1-9}$$

可见从错误模型算出的P_k、K_k均随着$K\to\infty$而趋于零。

因为
$$\begin{aligned}\hat{x}_k&=(1-K_k)\hat{x}_{k-1}+K_kz_k\\&=\frac{\sigma^2+(k-1)P_0}{\sigma^2+kP_0}\hat{x}_{k-1}+\frac{P_0}{\sigma^2+kP_0}z_k\end{aligned} \tag{10-1-10}$$

应用递推法，可以得出

$$\hat{x}_k=\frac{\sigma^2}{\sigma^2+kP_0}\hat{x}_0+\frac{P_0}{\sigma^2+kP_0}\sum_{i=1}^{k}z_i \tag{10-1-11}$$

不失一般性，取$T=1$，由测量方程得

$$z_k=x_0+k_a=v_k \tag{10-1-12}$$

代入(10-1-11)式中，得

$$\begin{aligned}\hat{x}_k&=\frac{\sigma^2}{\sigma^2+kP_0}\hat{x}_0+\frac{P_0}{\sigma^2+kP_0}\sum_{i=1}^{k}(x_0+ia+v_i)\\&=\frac{\sigma^2}{\sigma^2+kP_0}\hat{x}_0+\frac{P_0}{\sigma^2+kP_0}\left(kx_0+\frac{k(k+1)}{2}a+\sum_{i=1}^{k}v_i\right)\end{aligned} \tag{10-1-13}$$

气球的真实高度是

$$x_k=x_1+ka \tag{10-1-14}$$

估计误差为

$$\begin{aligned}\tilde{x}_k&=x_k-\hat{x}_k\\&=x_0+ka-\frac{\sigma^2}{\sigma^2+kP_0}\hat{x}_0\\&\quad-\frac{R_0}{\sigma^2+kP_0}\left(kx_0+\frac{k(k+1)}{2}a+\sum_{i=1}^{k}v_i\right)\\&=\frac{\sigma^2}{\sigma^2+kP_0}(x_0-\hat{x}_0)+\frac{2\sigma^2+kP_0-P_0}{2(\sigma^2+kP_0)}ka-\frac{P_0}{\sigma^2+kP_0}\sum_{i=1}^{k}v_i\end{aligned} \tag{10-1-15}$$

设测量噪声是零均值的和初值$\hat{x}_0$是无偏的，则有

$$E[\tilde{x}_k]=\frac{2\sigma^2+kP_0-P_0}{2(\sigma^2+kP_0)}ka \qquad (10\text{-}1\text{-}16)$$

显见，由错误模型得出的估计 $\hat{x}_k$ 是有偏的，而且偏量随 k 的增大而增大，其均方误差 $E[\tilde{x}_k^2]$ 亦随 k 的增大而增大。

上述现象称为滤波器的模型发散，收敛的和发散的滤波器的滤波结果如图 10-1-1 所示。

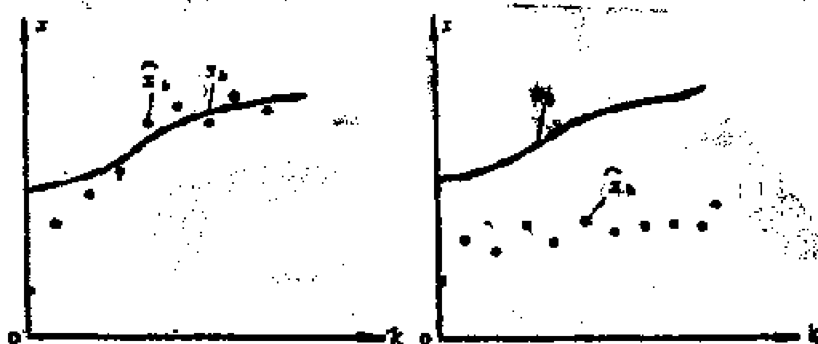


图10-1-1 收敛的与发散的滤波结果

造成此滤波器发散的原因显然是由于状态方程不准确所致。仔细考虑发散滤波器的机理，由滤波器的增益公式和方差公式

$$K_k=\frac{P_k}{\sigma^2},\quad P_k=\frac{P_0\sigma^2}{\sigma^2+kP_0}$$

可以看到增益 K_k 是随着时间的增大而递减的。当 K_k 很小时，测量值 z_k 已不起校准作用，滤波器于是按照错误的状态方程去作出估计，其结果亦必然是错误的，这种现象称为滤波器饱和，好象滤波器已不再能容入新的测量数据了。为了防止滤波器饱和，研究出了一些防止方法，下面分别讨论之。

（一）伪噪声引入法

由以上讨论可知，滤波器的饱和是由于增益变得太小所致。如果人为地在状态方程中加入噪声，就可以防止增益变得过小。

例10-1-2 令上例中的错误模型为

$$x_{k+1}=x_k+w_k \qquad (10\text{-}1\text{-}17)$$

$$z_{k+1}=x_{k+1}+v_{k+1} \qquad (10\text{-}1\text{-}18)$$

其中 w_k 是人为设定的噪声，并令 w_k 为零均值，方差等于 σ_w^2 的白噪声，求状态估计。

解：

$$\frac{1}{P_k}=\frac{1}{P_{k-1}+\sigma_w^2}+\frac{1}{\sigma^2}$$

或

$$P_k=\frac{\sigma^2(P_{k-1}+\sigma_w^2)}{P_{k-1}+\sigma_w^2+\sigma^2} \qquad (10\text{-}1\text{-}19)$$

和

$$K_k=\frac{P_k}{R_k}=\frac{P_{k-1}+\sigma_w^2}{P_{k-1}+\sigma_w^2+\sigma^2} \qquad (10\text{-}1\text{-}20)$$

由(10-1-20)式可知，尽管时间k在增长，K_k是不会趋于零的，故只要σ_ξ^2足够大，滤波器是不会饱和的。当然，由于人为地引入了状态噪声，将使滤波估计误差的方差增大。

上面仅研究了一个特例，再来研究一般情况。设引入的状态伪噪声的方差为Q_k，则有

$$P_k=(1-K_kH_k)(\Phi_{k/k-1}P_{k-1}\Phi_{k/k-1}^T+Q_{k-1}) \tag{10-1-21}$$

不失一般性，令$\Phi_{k/k-1}=\Phi$，$H_k=H$，并给定初始估计方差为P_0，则有

$$P_1=(1-K_1H)\Phi P_0\Phi^T+(1-K_1H)Q_0$$

$$\begin{aligned}P_2&=(1-K_2H)(\Phi P_1\Phi^T+Q_1)\\&=(1-K_2H)\Phi[(1-K_1H)\Phi P_0\Phi^T]\Phi^T\\&\quad+(1-K_2H)[\Phi(1-K_1H)Q_0\Phi^T+Q_1]\end{aligned}$$

依此类推，可得

$$\begin{aligned}P_k=&\left\{\prod_{i=k}^{1}(1-K_iH)\Phi\right\}P_0(\Phi^T)^k\\&+(1-K_kH)\left[\left\{\prod_{i=k-1}^{1}\Phi(1-K_iH)\right\}Q_1(\Phi^T)^{k-1}+\cdots\right.\\&+\left\{\prod_{i=k-1}^{k-2}\Phi(1-K_iH)\right\}Q_{k-2}(\Phi^T)^2\\&\left.+\{\Phi(1-K_{k-1}H)\}Q_{k-1}\Phi^T+Q_{k-1}\right]\end{aligned} \tag{10-1-22}$$

当$k\to\infty$时，上式中的第一项趋于零。但由于伪噪声Q_k的引入，P_k不会趋于零，从而增益$K_k=P_kH_k^TR_k^{-1}$亦不会趋于零，所以滤波器不会饱和。

因此，当我们缺乏对状态噪声的统计知识时，可以引入适当大小的伪噪声以保证滤波器是收敛的。引入的Q_k过小，不足以抑制滤波器发散，但引入的Q_k过大，则滤波器估计的方差增大。引入Q_k的大小可通过试验来确定。

（二）限定增益法

由以上讨论可知，滤波器的饱和是由于增益变得太小所致，这时新测量值已不起作用。如果人为地限定增益的下限，就可以防止滤波器饱和。

我们规定滤波器增益的下限为$\overline{K}$，当$K_k\geqslant\overline{K}$时，滤波器增益仍为

$$K_{k+1}=P_{k+1/k}H_{k+1}^T[H_{k+1}P_{k+1/k}H_{k+1}^T+R_{k+1}]^{-1} \tag{10-1-23}$$

故

$$\hat{X}_{k+1}=\hat{X}_{k+1/k}+K_{k+1}[Z_{k+1}-H_{k+1}\hat{X}_{k+1/k}] \tag{10-1-24}$$

当$K_k<\overline{K}$时，滤波估计取

$$\hat{X}_{k+1}=\hat{X}_{k+1/k}+\overline{K}[Z_{k+1}-H_{k+1}\hat{X}_{k+1/k}] \tag{10-1-25}$$

可见，此时卡尔曼滤波器退化成恒增益递归滤波器。$\overline{K}$可按照卡尔曼滤波的稳定性定理二选定，$\overline{K}$应满足使$(1-\overline{K}H_k)\Phi_{k/k-1}$的特征值位于单位圆内。

（三）渐消记忆滤波法

从第七章中我们已经知道，卡尔曼滤波器本身具有渐消记忆的功能，即把老的测量数据逐渐消除，而输出主要由较新数据加权组合而成的数据。如果人为地加强这一渐消记忆功能，就可以克服滤波器的模型发散。因为模型发散主要是由于状态方程设置不准确造成的，我们认为测量方程是正确的，所以，当滤波器的输出主要由最新的若干测量数据加权

组合而成时，那么滤波器输出就不会偏离真值太远，从而克服了发散现象。

由卡尔曼滤波公式

$$P_{k/k-1}=\Phi_{k/k-1}P_{k-1}\Phi_{k/k-1}^{T}+Q_{k-1} \tag{10-1-26}$$

$$P_{k}^{-1}=P_{k/k-1}^{-1}+H_{k}^{T}R_{k}^{-1}H_{k} \tag{10-1-27}$$

$$K_{k}=P_{k}H_{k}^{T}R_{k}^{-1} \tag{10-1-28}$$

可以看出，如果令P_{k-1}人为地增大一个因子α，即令

$$P_{k-1}\longrightarrow \alpha P_{k-1}=P_{k-1}^{*},\quad \alpha>1 \tag{10-1-29}$$

则得出滤波方程为

$$P_{k/k-1}^{*}=\Phi_{k/k-1}P_{k-1}^{*}\Phi_{k/k-1}^{T}+Q_{k-1} \tag{10-1-30}$$

$$P_{k}^{*-1}=P_{k/k-1}^{*-1}+H_{k}^{T}R_{k}^{-1}H_{k} \tag{10-1-31}$$

$$K_{k}^{*}=P_{k}^{*}H_{k}^{T}R_{k}^{-1} \tag{10-1-32}$$

显见，K_{k}^{*}将大于原来的K_{k}，由滤波器方程

$$\hat{X}_{k}^{*}=\hat{X}_{k/k-1}^{*}+K_{k}^{*}[Z_{k}-H_{k}\hat{Z}_{k/k-1}^{*}] \tag{10-1-33}$$

可以看到，由于增益K_{k}^{*}的增大，测量值Z_{k}的作用加强了。

由此得到渐消记忆卡尔曼滤波器如下：

$$\left.\begin{aligned}
&\hat{X}_{k+1/k}=\Phi_{k+1/k}\hat{X}_{k}\\
&P_{k+1/k}=\alpha\Phi_{k+1/k}P_{k}\Phi_{k+1/k}^{T}+\Gamma_{k}Q_{k}\Gamma_{k}^{T},\ \alpha>1\\
&K_{k+1}=P_{k+1/k}H_{k+1}^{T}[H_{k+1}P_{k+1/k}H_{k+1}^{T}+R_{k+1}]^{-1}\\
&\hat{X}_{k+1}=\hat{X}_{k+1/k}+K_{k+1}[Z_{k+1}-H_{k+1}\hat{X}_{k+1/k}]\\
&P_{k+1}=(1-K_{k+1}H_{k+1})P_{k+1/k}
\end{aligned}\right\} \tag{10-1-34}$$

（四）固定记忆滤波

固定记忆滤波的含义是在求t_k时刻的估计$\hat{X}_k$时，限定只利用最新的N个测量值

$$Z_{k-N-1},\ Z_{k-N+2},\ \cdots,Z_{k-1},Z_{k}$$

而把以前的老数据全部甩掉。由于只利用最新的N个数据采样长度，在这段时间内，模型不正确的影响不会大量积累。固定记忆滤波特别适用于统计性质阶段性突变的情况。

我们采用无状态噪声模型加以说明。

$$X_{k}=\Phi_{k/k-1}X_{k-1}+B_{k-1}U_{k-1} \tag{10-1-35}$$

$$Z_{k}=H_{k}X_{k}+V_{k} \tag{10-1-36}$$

由前面讨论可后，由上述模型构造的滤波器可能发散，现在采用下述方法来抑制发散。

从t_0时刻起应用常规卡尔曼滤波器递推N步，得到t_N时刻的估计$\hat{X}(N)$及其误差方差阵$P(N)$，并把$\hat{X}(N)$和$P(N)$存在存储器中。当滤波器继续运行至t_{2N}时刻，求出$\hat{X}(N)$对$\hat{X}(2N)$的最优预报及预报方差

$$\hat{X}(2N/N)=\Phi_{2N/N}\hat{X}(N)+\sum_{i=N+1}^{2N}\Phi_{2N/i}B_{i-1}U_{i-1} \tag{10-1-37}$$

$$P(_{2N/N})=\Phi_{2N/N}P_{N}\Phi_{2N/N}^{T} \tag{10-1-38}$$

用符号$\hat{X}_N(2N)$和$P_N(2N)$表示在t_{2N}时刻的限定记忆估计及其方差阵，于是有

$$\hat{X}_{N}(2N)=P_{N}(2N)[P^{-1}(2N)\hat{X}(2N)-P^{-1}(2N/N)\hat{X}(2N/N)] \tag{10-1-39}$$

$$P_N^{-1}(2N)=P^{-1}(2N)-P^{-1}(2N/N) \tag{10-1-40}$$

其中$\hat{X}(2N)$和$P(2N)$乃是常规滤波器在t_{2N}时刻的估计及其方差阵。上式的含义是把$\hat{X}(2N)$看成是$\hat{X}_N(2N)$与$\hat{X}(2N/N)$的最优加权最小二乘估计。

现在再把$\hat{X}_N(2N)$和$P_N(2N)$存储起来供以后应用，并作为常规滤波器的初值，常规滤波器运行至t_{3N}时刻，得出$\hat{X}(3N)$和$P(3N)$。在t_{3N}时刻，再作出预报$\hat{X}(3N/2N)$及其方差阵$P(3N/2N)$。

$$\hat{X}(3N/2N)=\Phi_{3N/2N}\hat{X}(2N)+\sum_{j=2N+1}^{3N}\Phi_{3N/2N}B_{j-1}U_{j-1} \tag{10-1-41}$$

$$P(3N/2N)=\Phi_{3N/2N}P(2_N)\Phi^T_{3N/2N} \tag{10-1-42}$$

同理，可得到t_{3N}时刻的限定记忆估计及其方差阵。

$$\hat{X}_N(3N)=P_N(3N)[P^{-1}(3N)\hat{X}(3N)-P^{-1}(3N/2N)\hat{X}(3N/2N)] \tag{10-1-43}$$

$$P_N^{-1}(3N)=P^{-1}(3N)-P^{-1}(3N/2N) \tag{10-1-44}$$

依此类推，就得到各个t_k时刻的估计。

必须指出，上述四种抑制滤波器模型发散的方法都是以牺牲滤波精度为代价的，属于消极措施，积极的方法应当是建立正确的信号模型。当对于模型的先验知识不足时，应当采用自适应方法从观测值来建立模型与估计模型参数。为此，近年来自适应滤波理论愈益受到重视。

由于自适应卡尔曼滤波的理论和实践都较新，理论上的困难亦较大，尚没有完整统一的理论，这里仅介绍几种比较有效的实用方法。

§10-2 机动量的检测与估计

从上面的讨论可知，造成模型发散的根本原因在于所用的状态方程不正确。当真实状态变化规律发生变动时，状态方程应适时地作出调整。例如，在k时刻之前，目标基本上按等速度航行，状态方程亦按二维设定，所以在k时刻之前，滤波器能正常工作；但在k时刻之后，目标进行大的加速，若滤波器模型未作调整，滤波器就可能出现发散。因此，为了防止滤波器发散和得到好的估计效果，我们希望能够及时发现目标进行加速的时刻，并估计出加速度的大小，以便及时调整信号模型。

设目标未作机动前的状态方程为

$$\begin{bmatrix} x_{k+1} \\ \dot{x}_{k+1} \end{bmatrix}=\begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_k \\ \dot{x}_k \end{bmatrix}+\begin{bmatrix} 0 \\ 1 \end{bmatrix}w_k \tag{10-2-1}$$

当目标有机动加速度u_k时，状态方程应为

$$\begin{pmatrix} x_{k+1} \\ \dot{x}_{k+1} \end{pmatrix}=\begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x_k \\ \dot{x}_k \end{pmatrix}+\begin{pmatrix} \frac{T^2}{2} \\ T \end{pmatrix}u_k+\begin{pmatrix} 0 \\ 1 \end{pmatrix}w_k \tag{10-2-2}$$

或

$$X_{k+1}=\Phi X_k+BU_k+\Gamma W_k$$

其中W_k为零均值速度扰动。由于大气湍流、阵风、发动机推力起伏等因素影响，可看成

是零均值白噪声。

测量值Z_k是距离，故测量方程为

$$Z_k=[1,\ 0]\begin{bmatrix}x_k\\ \dot{x}_k\end{bmatrix}+V_k \tag{10-2-3}$$

或

$$Z_k=HZ_k+V_k$$

偏量分析

对于有大的加速度的目标，如仍用（10-2-1）中的模型表示，相应的滤波器中将出现如下的问题：

1．在预报估计中将出现偏量，从而新息中亦出现偏量。

因为预报误差为

$$\begin{aligned}\tilde{X}_{k+1/k}&=X_{k+1}-\hat{X}_{k+1/k}\\ &=\begin{cases}\Phi\tilde{X}_k+\Gamma W_k & U_k=0\\ \Phi\tilde{X}_k+\Gamma W_k+BU_k & U_k\neq 0\end{cases}\end{aligned} \tag{10-2-4}$$

$$E[\tilde{X}_{k+1/k}]=\begin{cases}0 & U_k=0\\ BU_k & U_k\neq 0\end{cases} \tag{10-2-5}$$

而滤波器中新息为

$$\begin{aligned}\tilde{Z}_{k+1}&=Z_{k+1}-H\hat{X}_{k+1/k}\\ &=H\tilde{X}_{k+1/k}+V_{k-1}\end{aligned} \tag{10-2-6}$$

$$E[\tilde{Z}_{k+1}]=\begin{cases}0 & U_k=0\\ HBU_k & U_k\neq 0\end{cases} \tag{10-2-7}$$

2．预报误差的方差增大

$$\begin{aligned}P_{k+1/k}&=E[\tilde{X}_{k+1/k}\tilde{X}^T_{k+1/k}]\\ &=\begin{cases}\Phi P_k\Phi^T+\Gamma Q_k\Gamma^T & U_k=0\\ \Phi P_k\Phi^T+\Gamma Q_k\Gamma^T+BU_k^2B^T & U_k\neq 0\end{cases}\end{aligned} \tag{10-2-8}$$

由以上分析可见，如果实际存在U_k而滤波器仍按$U_k=0$工作，新息中将出现偏量，方差增大，将使滤波性能变坏甚至发散。为此必须及时对滤波器修正，这需要检测U_k的存在并估计其大小。

机动检测

假定在U_k加上之前，滤波器已进入稳态，这时新息为零均值白噪声，其方差为

$$P_{\tilde{z}}=HP_{k+1/k}H^T+R_{k+1} \tag{10-2-9}$$

而在加上U_k之后，新息出现偏量，且方差增大。所以，在U_k加上的前后，新息的概率分布将发生变化，如图10-2-1所示。

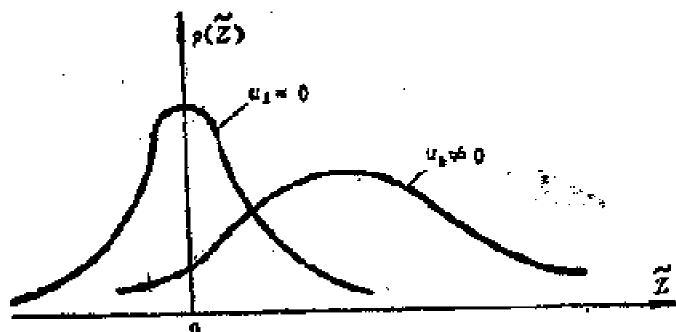


图10-2-1　U_k加上前后$\tilde{Z}_k$的概率分布

因此，我们可以利用$|\tilde{Z}_k|$作为检验统计量来检测U_k的存在，这类似于从零均值高斯噪

声中检测直流偏量的问题，利用符号法检测原理，看连续几次满足下式

$$|\tilde{Z}| \geq b\sqrt{P_{\tilde{z}}} \tag{10-2-10}$$

其中b是某正常数，$P_{\tilde{z}}$如(10-2-9)式所示。

若满足(10-2-10)式，则判为 $U_k \neq 0$，否则判为 $U_k = 0$。显然，b 增大，虚警概率降低，但检测概率也降低，反之则增大。一旦判决出$U_k \neq 0$，就检测到了U_k的存在和加上U_k的时刻。

估计U_k的大小

当检测到$U_k \neq 0$之后，希望对它的大小作出估计。

设在t_k时刻加上U_k，对于按$U_k=0$工作的滤波器来说，在t_k之前的预报是正确的和无偏的，而从$(k+1)T$时刻开始，预报出现偏差。

记无偏预报为$\hat{X}_{k+1/k}$，有偏预报为$\check{X}_{k+1/k}$。按照公式$\hat{X}_{k+1/k} = \Phi\hat{X}_k + BU_k$作出的预报才是无偏的。我们列出二次有偏预报$\check{X}_{k+1/k}$和$\check{X}_{k+2/k+1}$。

$$\begin{aligned}\check{X}_{k+1/k} &= \Phi\hat{X}_k \\ &= \Phi(I-K_kH)\hat{X}_{k/k-1} + \Phi K_k Z_k \end{aligned}\tag{10-2-11}$$

$$\begin{aligned}\check{X}_{k+2/k+1} &= \Phi\hat{X}_{k+1} \\ &= \Phi\{(I-K_{k+1}H)\hat{X}_{k+1/k} + K_{k+1}Z_{k+1}\} \\ &= \Phi(I-K_{k+1}H)[\Phi(I-K_kH)\hat{X}_{k/k-1} + \Phi K_k Z_k] + \Phi K_{k+1} Z_{k+1}\end{aligned}\tag{10-2-12}$$

由此得有偏新息为

$$\check{\tilde{Z}}_{k+1/k} = Z_{k+1} - H\check{X}_{k+1/k} \tag{10-2-13}$$

$$\check{\tilde{Z}}_{k+2/k+1} = Z_{k+2} - H\check{X}_{k+2/k+1} \tag{10-2-14}$$

由正确的预报公式得出无偏预报为$\hat{Z}_{k+1/k}$和$\hat{Z}_{k+2/k+1}$。

$$\begin{aligned}\hat{X}_{k+1/k} &= \Phi\hat{X}_k + BU_k \\ &= \Phi(I+K_kH)\hat{X}_{k/k-1} + \Phi K_k Z_k + BU_k\end{aligned}\tag{10-2-15}$$

$$\begin{aligned}\hat{X}_{k+2/k+1} &= \Phi\hat{X}_{k+1} + BU_{k+1} \\ &= \Phi[(I-K_{k+1}H)\hat{X}_{k+1/k} + K_{k+1}Z_{k+1}] + BU_{k+1} \\ &= \Phi\{(I-K_{k+1}H)[\Phi(I-K_kH)\hat{X}_{k/k-1} + \Phi K_k Z_k + BU_k] \\ &\quad + K_{k+1}Z_{k+1}\} + BU_{k+1}\end{aligned}\tag{10-2-16}$$

由此算出无偏新息为

$$\tilde{Z}_{k+1/k} = Z_{k+1} - H\hat{X}_{k+1/k} \tag{10-2-17}$$

$$\tilde{Z}_{k+2/k+1} = Z_{k+2} - H\hat{X}_{k+2/k+1} \tag{10-2-18}$$

再求两种新息之差为

$$\check{\tilde{Z}}_{k+1/k} - \tilde{Z}_{k+1/k} = HBU_k \tag{10-2-19}$$

$$\check{\tilde{Z}}_{k+2/k+1} - \tilde{Z}_{k+2/k+1} = HBU_{k+1} + H\Phi(I-K_{k+1}H)BU_k \tag{10-2-20}$$

改写成矩阵形式，得

$$\begin{pmatrix} \breve{\tilde{Z}}_{k+1/k} \\ \breve{\tilde{Z}}_{k+2/k+1} \end{pmatrix} = \begin{pmatrix} HBU_k \\ HBU_{k+1} + H\Phi(1 - K_{k+1}H)BU_k \end{pmatrix} + \begin{pmatrix} \tilde{Z}_{k+1/k} \\ \tilde{Z}_{k+2/k+1} \end{pmatrix} \quad (10\text{-}2\text{-}21)$$

通常机动时间比采样周期T大得多，可以认为外加控制加速度$U_{k-1} = U_k = U$，把(10-2-21)式写成如下形式

$$\breve{\tilde{Z}} = CU + \tilde{Z}_e \quad (10\text{-}2\text{-}22)$$

其中 $$C = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} HB \\ HB + H\Phi(1 - K_{k+1}H)B \end{pmatrix} \quad (10\text{-}2\text{-}23)$$

因为无偏估计时，新息为白噪声，故有

$$P_{\tilde{z}_e} = E\left\{\begin{pmatrix} \tilde{Z}_{k+1/k} \\ \tilde{Z}_{k+2/k+1} \end{pmatrix}[\tilde{Z}_{k+1/k}, \ \tilde{Z}_{k+2/k+1}]\right\}$$

$$= \begin{pmatrix} P_{\tilde{z}_{k+1/k}} & 0 \\ 0 & P_{\tilde{z}_{k+2/k+1}} \end{pmatrix} \quad (10\text{-}2\text{-}24)$$

因此，可以得出U的最优加权最小二乘估计为

$$\hat{U} = [C^T P_{\tilde{z}_e}^{-1} C]^{-1} C^T P_{\tilde{z}_e}^{-1} \breve{\tilde{Z}}$$

$$= \sigma_u^2 C^T P_{\tilde{z}_e}^{-1} \breve{\tilde{Z}} \quad (10\text{-}2\text{-}25)$$

其中σ_u^2为估计误差方差

$$\sigma_u^2 = [C^T P_{\tilde{z}_e}^{-1} C]^{-1} \quad (10\text{-}2\text{-}26)$$

因为$H = [1, 0]$，由(10-2-23)式可知，c_1和c_2都是标量，故得出

$$\sigma_u^2 = [c_1^2 P_{\tilde{z}_{k+1/k}}^{-1} + c_2^2 P_{\tilde{z}_{k+2/k+1}}^{-1}]^{-1} \quad (10\text{-}2\text{-}27)$$

$$\hat{u} = \sigma_u^2 [c_1, \ c_2] \begin{pmatrix} P_{\tilde{z}_{k+1/k}}^{-1} & 0 \\ 0 & P_{\tilde{z}_{k+2/k}}^{-1} \end{pmatrix} \begin{pmatrix} \breve{\tilde{z}}_{k+1/k} \\ \breve{\tilde{z}}_{k+2/k+1} \end{pmatrix}$$

$$= \sigma_u^2 [c_1 P_{\tilde{z}_{k+1/k}}^{-1} \breve{\tilde{z}}_{k+1/k} + c_2 P_{\tilde{z}_{k+2/k+1}}^{-1} \breve{\tilde{z}}_{k+2/k+1}] \quad (10\text{-}2\text{-}28)$$

在得到加速度估计$\hat{u}$之后，把它代入预报方程中，便得到新的最优滤波器。

因为加速度通常只有一定持续时间，当目标加速度消失时，检验统计量$|\tilde{z}|$会再次超过门限，可以用$\hat{u} = 0$试探，如果$|\tilde{z}|$不超过门限，即认为目标已没有加速度。

上述方法的不足之处在于加速度的检测与估计需要一定时间，因而对滤波器的修正有

滞后时间，在此滞后期间，滤波效果变坏。这里为了简明起见，只用二个新息估计 u，而为了较好地估计 u，应当应用多个新息。

克服上述缺点的方法是把加速度看成是状态变量，构造三维卡尔曼滤波器，对目标的距离、速度、加速度进行递推估计。这样做法显然加大了计算量。对于目标大部分时间按等速航行并且数据率很高的情况（高数据率有助于提高滤波精度），采用二维跟踪滤波器加机动检测与估计器的方法是比较有效的。

§10-3　最优稳态增益的估计

在第七章中曾经指出，对于定常信号模型而言，存在一个最佳稳态增益$\overline{K}$，$\overline{K}$决定于信号模型参数，但由于噪声方差Q、R为未知而无法求出。米莱（Mehra）提出，可以根据新息的相关序列对$\overline{K}$进行自适应估计。

在第七章中已指出，稳态最优滤波器的新息$r_k = Z_k - H\hat{X}_{k/k-1}$是白噪声序列，其相关函数为

$$
\begin{aligned}
A_i &= E[r_k r_{k-i}^T] \\
&= \begin{cases} H\overline{P^-}H^T + R & i=0 \\ 0 & i \neq 0 \end{cases}
\end{aligned}
\tag{10-3-1}
$$

其中$\overline{P^-}$为稳态预报误差方差。$\overline{P^-}$应满足下列稳态黎卡蒂方程。

$$
\overline{P^-} = \Phi(\overline{P^-} - \overline{K}H\overline{P^-})\Phi^T + \Gamma Q \Gamma^T \tag{10-3-2}
$$

并且最佳稳态增益$\overline{K}$满足下述方程

$$
\overline{K} = \overline{P^-}H^T[H\overline{P^-}H^T + R]^{-1} \tag{10-3-3}
$$

现在对同样的信号模型构造一个恒增益递归滤波器，令增益为K_s，$K_s \neq \overline{K}$。显然，此时滤波器是非最优的。考察此滤波器中的新息及新息的相关函数。

$$
\begin{aligned}
r_k &= Z_k - H\hat{X}_{k/k-1} = H(X_k - \hat{X}_{k/k-1}) + V_k \\
&= He_{k/k-1} + V_k
\end{aligned}
\tag{10-3-4}
$$

$$
\begin{aligned}
A_i &= E[r_k r_{k-i}^T] \\
&= \begin{cases} HE[e_k e_{k-i}^T]H^T + HE[e_k V_{k-i}^T], & i>0 \\ HP_{k/k-1}H^T + R, & i=0 \end{cases}
\end{aligned}
\tag{10-3-5}
$$

把预报误差$e_k = X_k - \hat{X}_{k/k-1}$表达为

$$
e_k = \Phi(1 - K_S H)e_{k-1} - \Phi K_S V_{k-1} + \Gamma W_{k-1} \tag{10-3-6}
$$

其中　$e_{k-1} = X_{k-1} - \hat{X}_{k-1/k-2}$ (10-3-7)

用代入法求出e_{k-i}，再求出稳态预报误差的相关函数，

$$
E[e_k e_{k-i}^T] = [\Phi(1 - K_s H)]^i P^- \tag{10-3-8}
$$

其中P^-为次优稳态预报误差方差阵，P^-满足下述方程

$$
P^- = \Phi[(1 - K_s H)P^-(1 - K_S H)^T + K_s R K_s^T]\Phi^T + \Gamma Q \Gamma^T \tag{10-3-9}
$$

上式中的Q、R应为准确值。

同理可求出

$$E[e_k V_{k-i}^T] = -[\Phi(I-K_s H)]^{i-1}\Phi K_s R \tag{10-3-10}$$

把(10-3-8)、(10-3-10)式代入A_i表达式中，得

$$A_i = \begin{cases} H[\Phi(I-K_s H)]^{i-1}\Phi[P^- H^T - K_s A_0] \\ HP^- H^T + R \end{cases} \tag{10-3-11}$$

相关函数A_i可用时间平均求得其估计

$$\hat{A}_i = \frac{1}{N}\sum_{j=k-N+1}^{k} r_j r_{j-i}^T \tag{10-3-12}$$

把得出的$\hat{A}_i$值代入(10-3-11)式中，再求$P^- H^T$的估计。各相关函数如下：

$i=1$，$\hat{A}_1 = H\Phi(\widehat{P^- H^T} - K_s \hat{A}_0)$

或 $H\Phi(\widehat{P^- H^T}) = \hat{A}_1 + H\Phi K_s \hat{A}_0$

$i=2$，$\hat{A}_2 = H\Phi(I-K_s H)\Phi(\widehat{P^- H^T} - K_s \hat{A}_0)$

或 $H\Phi^2(\widehat{P^- H^T}) = \hat{A}_2 + H\Phi K_s \hat{A}_1 + H\Phi^2 K_s \hat{A}_0$

……………………

$i=N$，$\hat{A}_N = H[\Phi(I-K_s H)]^{N-1}\Phi(\widehat{P^- H^T} - K_s \hat{A}_0)$

或 $H\Phi^N(\widehat{P^- H^T}) = \hat{A}_N + H\Phi K_s \hat{A}_{N-1} + \cdots + H\Phi^N K_s \hat{A}_0$

记

$$M = \begin{pmatrix} H \\ H\Phi \\ \vdots \\ H\Phi^{N-1} \end{pmatrix}\Phi \tag{10-3-13}$$

则有

$$M(\widehat{P^- H^T}) = \begin{pmatrix} \hat{A}_1 + H\Phi K_s \hat{A}_0 \\ \hat{A}_2 + H\Phi K_s \hat{A}_1 + H\Phi^2 K_s \hat{A}_0 \\ \cdots\cdots \\ \hat{A}_N + H\Phi K_s \hat{A}_{N-1} + \cdots + H\Phi^N K_s \hat{A}_0 \end{pmatrix} \tag{10-3-14}$$

于是得到估计$\widehat{P^- H^T}$为

$$\widehat{P^- H^T} = (M_T M)^{-1} M \begin{pmatrix} \hat{A}_1 + H\Phi K_s \hat{A}_0 \\ \hat{A}_2 + H\Phi K_s \hat{A}_1 + H\Phi^2 K_s \hat{A}_0 \\ \cdots\cdots \\ \hat{A}_N + H\Phi K_s \hat{A}_{N-1} + \cdots + H\Phi^N K_s \hat{A}_0 \end{pmatrix} \tag{10-3-15}$$

其中$(M^T M)^{-1} M$称为M的伪逆。

由此还可以得到R的估计为

$$\hat{R} = \hat{A}_0 - H\widehat{P^- H^T} \tag{10-3-16}$$

因此次优增益K_s时的预报误差方差阵为P^-，P^-与K_s的关系如（10-3-15）所示。再设在最优稳态增益$\overline{K}$时，预报误差方差阵为$\overline{P^-}$，$\overline{P^-}$与$\overline{K}$满足下列稳态方程，

$$\overline{P}^- = \Phi(\overline{P}^- - \overline{K}H\overline{P}^-)\Phi^T + \Gamma Q\Gamma^T \tag{10-3-17}$$

记 $\delta P^- = \overline{P}^- - P^-$

令(10-3-17)式减去(10-3-9)式，得

$$\delta P^- = \Phi[\delta P^- - KHP^- + P^-H^TK_s - K_sRK_s^T - \overline{K}H\overline{P}^- - K_sHP^-H^TK_s^T]\Phi^T \tag{10-3-18}$$

因为

$$\overline{K}H\overline{P}^- = \overline{P}^-H^T[H\overline{P}^-H^T + R]^{-1}H\overline{P}^-$$

$$\overline{P}^-H^T = (P^- + \delta P^-)H^T$$

$$K_SRK_S^T + K_SHP^-H^TK_S^T = K_S(R + HP^-H^T)K_S^T$$

$$\hat{R} = \hat{A}_0 - HP^-H^T$$

把以上关系代入（10-3-18）式中，得

$$\delta P^- = \Phi[\delta P^- - (P^-H + \delta P^-H^T)(A_0 + H\delta P^-H^T)^{-1}(HP^- + H\delta P^-) + K_sHP^- + P^-HK_s^T + K_sA_0K_s^T]\Phi^T \tag{10-3-19}$$

把已算出的$\hat{A}_0$和$\widehat{P^-H^T}$代入上式就可解出$\widehat{\delta P^-}$。

因为最优稳态增益是

$$\begin{aligned}\overline{K} &= \overline{P}^-H^T[H\overline{P}^-H^T + R]^{-1}\\ &= (P^- + \delta P^-)H^T[HP^-H^T + H\delta P^-H^T + R]^{-1}\end{aligned} \tag{10-3-20}$$

把$\widehat{\delta P^-}$和$\widehat{P^-H^T}$代入，便得到最优稳态增益估计为

$$\hat{\overline{K}} \approx (\widehat{P^-H^T} + \widehat{\delta P^-}H^T)(\hat{A}_0 + H\widehat{\delta P^-}H^T)^{-1} \tag{10-3-21}$$

在非实时运算时，可以把$\hat{\overline{K}}$看成是新的非最优增益K_s。重复以上步骤，得出新的最优增益估计，如此反复计算，就得到最优增益的精确估计，随着迭代步数增长，滤波器中的新息将愈来愈趋于白色。

根据(10-3-17)式，有

$$\Gamma Q\Gamma^T = \overline{P}^- - \Phi(\overline{P}^- - \overline{K}H\overline{P}^-)\Phi^T \tag{10-3-22}$$

其中$\overline{P}^-$可由下式算出其估计

$$\hat{\overline{P}}^- = \hat{\overline{K}}\hat{A}_0(H^T)^* = \hat{\overline{K}}\hat{A}_0(H^TH)^{-1}H \tag{10-3-23}$$

于是可求出状态噪声方差阵的估计为

$$\widehat{\Gamma Q\Gamma^T} = \hat{\overline{P}}^- - \Phi(\hat{\overline{P}}^- - \hat{\overline{K}}H\hat{\overline{P}}^-)\Phi^T \tag{10-3-24}$$

§10-4 噪声统计量的差分法估计

设有定常二阶模型为

$$\left.\begin{aligned}\begin{pmatrix}x_k\\ \dot{x}_k\end{pmatrix} &= \begin{pmatrix}1 & T\\ 0 & 1\end{pmatrix}\begin{pmatrix}x_{k-1}\\ \dot{x}_{k-1}\end{pmatrix} + \begin{pmatrix}0\\ 1\end{pmatrix}u_{k-1}\\ z_k &= x_k + v_k\end{aligned}\right\} \tag{10-4-1}$$

其中w_k、v_k为独立零均值白噪声，其方差Q、R为未知，现在要由z_k序列求出估计值$\hat{Q}$、$\hat{R}$。在得到估计$\hat{Q}$、$\hat{R}$之后，就可以用卡尔曼滤波器求状态估计$\hat{x}_k$、$\hat{\dot{x}}_k$。

显见，测量序列z_k是非平稳的，但可以证明二阶差分$\nabla^2 z_k$却是平稳序列，由差分定义得

$$\left.\begin{aligned}\nabla z_k &= z_k - z_{k-1}\\ \nabla^2 z_k &= (z_k - z_{k-1}) - (z_{k-1} - z_{k-2})\end{aligned}\right\} \tag{10-4-2}$$

而

$$\begin{aligned}z_k - z_{k-1} &= x_k + v_k - x_{k-1} - v_{k-1}\\ &= T\dot{x}_{k-1} + v_k - v_{k-1}\end{aligned} \tag{10-4-3}$$

$$z_{k-1} - z_{k-2} = T\dot{x}_{k-2} + v_{k-1} - v_{k-2} \tag{10-4-4}$$

故

$$\begin{aligned}\nabla^2 z_k &= T(\dot{x}_{k-1} - \dot{x}_{k-2}) + v_k - 2v_{k-1} + v_{k-2}\\ &= Tw_{k-2} + v_k - 2v_{k-1} + v_{k-2}\end{aligned} \tag{10-4-5}$$

显然，$\nabla^2 z_k$为噪声的线性组合，即二阶差分$\nabla^2 z_k$为平稳零均值有色噪声，其相关函数为

$$\begin{aligned}\Phi(0) &= E[(\nabla^2 z_k)^2]\\ &= T^2 Q + 6R\end{aligned} \tag{10-4-6}$$

$$\begin{aligned}\Phi(1) &= E[(\nabla^2 z_k)(\nabla^2 z_{k-1})]\\ &= E[(Tw_{k-2} + v_k - 2v_{k-1} + v_{k-2})(Tw_{k-3} + v_{k-1} - 2v_{k-2} + v_{k-3})]\\ &= -4R\end{aligned} \tag{10-4-7}$$

由于$\nabla^2 z_k$为平稳序列，根据各态历经性质，相关函数$\Phi(0)$、$\Phi(1)$可用时间平均求得其估计。相关函数$\Phi(0)$的估计为

$$\hat{\Phi}_N(0) = \frac{1}{N}\sum_{i=0}^{N-1}(\nabla^2 z_i)^2 \tag{10-4-8}$$

相关函数$\Phi(1)$的估计为

$$\hat{\Phi}_N(1) = \frac{1}{N}\sum_{i=1}^{N-1}(\nabla^2 z_i)(\nabla^2 z_{i-1}) \tag{10-4-9}$$

将$\hat{\Phi}_N(0)$、$\hat{\Phi}_N(1)$代入(10-4-6)、(10-4-7)式中，得

$$\hat{R} = -\frac{1}{4}\hat{\Phi}_N(1) \tag{10-4-10}$$

$$\hat{Q} = \frac{1}{T^2}\left(\hat{\Phi}_N(0) + \frac{3}{2}\hat{\Phi}_N'(1)\right) = \frac{1}{T^2}(\Phi(0) - 6\hat{R}) \tag{10-4-11}$$

相关函数$\hat{\Phi}_N(j)$的逆推算法，在这里取

$$\hat{\Phi}_N(j) = \frac{1}{N}\sum_{i=j}^{N-1}\nabla^2 z_i \nabla^2 z_{i-j}, \qquad j = 0, 1$$

虽然$\hat{\Phi}_N(j)$是有偏估计，但它给出最小均方误差，比对应的无偏估计要好，可以证明，$\hat{\Phi}_N(j)$是$\Phi(j)$的渐近无偏一致估计。因为

$$\Phi_{N+1}(j)=\frac{1}{N+1}\sum_{i=j}^{N}\nabla^2 z_i\nabla^2 z_{i-j}$$

$$=\frac{1}{N+1}\left[\sum_{i=j}^{N-1}\nabla^2 z_i\nabla^2 z_{i-j}+\nabla^2 z_N\nabla^2 z_{N-j}\right]$$

$$=\left(\frac{1}{N}-\frac{1}{N(N+1)}\right)\left[\sum_{i=j}^{N-1}\nabla^2 z_i\nabla^2 z_{i-j}+\nabla^2 z_N\nabla^2 _{N-j}\right]$$

$$=\hat{\Phi}_N(j)-\frac{1}{N+1}\Phi_N(j)+\frac{1}{N+1}\nabla^2 z_N\nabla^2 z_{N-j}$$

$$=\hat{\Phi}_N(j)+\frac{1}{N+1}\left[\nabla^2 z_N\nabla^2 z_{N-j}-\hat{\Phi}_N(j)\right]$$

上式是求相关函数的估计的递推算法，选取初值$\Phi_0(j)$，就可逐步计算出$\Phi_N(j)$，并求相应的$\hat{Q}$、$\hat{R}$直到满足要求的精度为止。

在求$\hat{\Phi}_N(0)$和$\hat{\Phi}_N(1)$时，样本数N应取得足够大，在此期间应保证w_k、v_k是平稳的，从而$\nabla^2 z_k$是平稳的。相应的卡尔曼滤波器为

$$\left.\begin{aligned}
&\hat{X}_{k+1/k}=\Phi\hat{X}_k\\
&P_{k+1/k}=\Phi P_k\Phi^T+\Gamma\hat{Q}\Gamma^T\\
&K_{k+1}=P_{k+1/k}H^T[HP_{k+1/k}H^T+\hat{R}]^{-1}\\
&\hat{X}_{k+1}=\hat{X}_{k+1/k}+K_{k+1}[z_k-H\hat{X}_{k+1/k}]\\
&P_{k+1}=[I-K_{k+1}H]P_{k+1/k}
\end{aligned}\right\}\tag{10-4-12}$$

显然在以上分析中，我们假定模型是定常的，噪声是平稳的。但是当噪声统计量Q、R变化缓慢时，可以根据有限个观测值来得到相关函数的估计，取k时刻的估计为

$$\hat{\Phi}_k(0)=\frac{1}{N}\sum_{i=k-N+1}^{k}(\nabla^2 z_i)^2\tag{10-4-13}$$

$$\hat{\Phi}_k(1)=\frac{1}{N}\sum_{i=k-N+1}^{k}(\nabla^2 z_i)(\nabla^2 z_{i-1})\tag{10-4-14}$$

令N保持不变，取$k+1$时刻的估计为

$$\hat{\Phi}_{k+1}(0)=\frac{1}{N}\sum_{i=k-N+2}^{k+1}(\nabla^2 z_i)^2\tag{10-4-15}$$

$$\hat{\Phi}_{k+1}(1)=\frac{1}{N}\sum_{i=k-N+2}^{k+1}(\nabla^2 z_i)(\nabla^2 z_{i-1})\tag{10-4-16}$$

对R_k、Q_k的估计为

$$\hat{R}_k=-\frac{1}{4}\hat{\Phi}_k(1)\tag{10-4-17}$$

$$\hat{Q}_k=\frac{1}{T^2}\left[\hat{\Phi}_k(0)+\frac{3}{2}\hat{\Phi}_k(1)\right]\tag{10-4-18}$$

在(10-4-12)式卡尔曼滤波器中用$\hat{R}_k$、$\hat{Q}_k$代替$\hat{R}$、$\hat{Q}$只要计算速度跟得上，就达到对时

变噪声统计模型的自适应估计。

W_k为有色噪声情况

若(10-4-1)式中二阶定常模型的噪声w_k为用一阶马尔科夫过程表示的有色噪声

$$w_k = \rho w_{k-1} + \xi_{k-1} \tag{10-4-19}$$

其中ξ_k为平稳零均值白噪声，其方差σ_ξ^2为未知由§8-2得知，$E[w_k^2]=Q$与σ_ξ^2的关系为

$$\sigma_\xi^2 = E[\xi_k^2] = (1-\rho^2)Q \tag{10-4-20}$$

根据第八章中状态矢量扩充法原理，立即得出上述情况下的信号模型为

$$\left.\begin{aligned} \begin{pmatrix} x_k \\ \dot{x}_k \\ w_k \end{pmatrix} &= \begin{pmatrix} 1 & T & 0 \\ 0 & 1 & 1 \\ 0 & 0 & \rho \end{pmatrix}\begin{pmatrix} x_{k-1} \\ \dot{x}_{k-1} \\ w_{k-1} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \xi_{k-1} \end{pmatrix} \\ z_k &= x_k + v_k \end{aligned}\right\} \tag{10-4-21}$$

其中未知参数为ρ、σ_ξ^2、R，现在用差分法求它们的估计。

求测量序列的二阶差分

$$\begin{aligned} \nabla^2 z_k &= Tw_{k-2} + v_k - 2v_{k-1} + v_{k-2} \\ &= T\rho w_{k-3} + T\xi_{k-3} + v_k - 2v_{k-1} + v_{k-2} \end{aligned} \tag{10-4-22}$$

显见，$\nabla^2 z_k$为零均值平稳有色噪声序列，其相关函数为

$$\Phi(0) = E[(\nabla^2 z_k)^2] = T^2\rho^2 Q + T^2(1-\rho^2)Q + 6R \tag{10-4-23}$$

$$\begin{aligned} \Phi(1) &= E[(\nabla^2 z_k)(\nabla^2 z_{k-1})] \\ &= \rho^2 T^2 E[w_{k-3}w_{k-4}] + \rho T^2 E[w_{k-3}\xi_{k-4}] - 4R \\ &= \rho^3 T^2 Q + \rho T^2(1-\rho^2)Q - 4R = \rho T^2 Q - 4R \end{aligned} \tag{10-4-24}$$

$$\begin{aligned} \Phi(2) &= E[(\nabla^2 z_k)(\nabla^2 z_{k-2})] \\ &= E[(T\rho w_{k-3} + T\xi_{k-3} + v_k - 2v_{k-1} + v_{k-2}) \\ &\quad \times (T\rho w_{k-5} + T\xi_{k-5} + v_{k-2} - 2v_{k-3} + v_{k-4})] \\ &= R \end{aligned} \tag{10-4-25}$$

再求得$\Phi(0)$、$\Phi(1)$、$\Phi(2)$的估计为

$$\left.\begin{aligned} \hat{\Phi}_N(0) &= \frac{1}{N}\sum_{i=k-N+1}^{k}(\nabla^2 z_i)^2 \\ \hat{\Phi}_N(1) &= \frac{1}{N}\sum_{i=k-N+1}^{k}(\nabla^2 z_i)(\nabla^2 z_{i-1}) \\ \hat{\Phi}_N(2) &= \frac{1}{N}\sum_{i=k-N-1}^{k}(\nabla^2 z_i)(\nabla^2 z_{i-2}) \end{aligned}\right\} \tag{10-4-26}$$

于是求得

$$\left.\begin{aligned} \hat{R} &= \hat{\Phi}_N(2) \\ \hat{Q} &= \frac{1}{T^2}[\hat{\Phi}_N(0) - 6\hat{\Phi}_N(2)] \\ \hat{\rho} &= \frac{1}{T^2\hat{Q}}[\hat{\Phi}_N(1) + 4\hat{\Phi}_N(2)] \end{aligned}\right\} \tag{10-4-27}$$

把$\hat{R}$、$\hat{Q}$、$\hat{\rho}$代入相应的卡尔曼滤波器中，就得到状态估计。

状态噪声不为零均值的情况

许多实际情况状态噪声不能视为零均值，仍以上述二阶定常模型为例，信号模型表达为

$$
\left.\begin{aligned}
&\begin{pmatrix} x_k \\ \dot{x}_k \end{pmatrix} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{k-1} \\ \dot{x}_{k-1} \end{pmatrix} + \begin{pmatrix} 0 \\ q + w_{k-1} \end{pmatrix} \\
&z_k = x_k + v_k
\end{aligned}\right\} \tag{10-4-28}
$$

其中q为状态噪声均值，w_{k-1}仍是平稳零均值白噪声，模型中的未知参数为q、Q和R。

首先用差分法求参数估计$\hat{q}$、$\hat{Q}$和$\hat{R}$。

由(10-4-5)式，可求出z_k序列的二阶差分

$$\nabla^2 z_k = Tq + Tw_{k-2} + v_k - 2v_{k-1} + v_{k-2} \tag{10-4-29}$$

$\nabla^2 z_k$为平稳有色噪声，求其相关函数

$$
\begin{aligned}
\Phi(0) &= E[(\nabla^2 z_k)^2] \\
&= T^2 q + T^2 Q + 6R
\end{aligned} \tag{10-4-30}
$$

$$
\begin{aligned}
\Phi(1) &= E[(\nabla^2 z_k)(\nabla^2 z_{k-1})] \\
&= T^2 q - 4R
\end{aligned} \tag{10-4-31}
$$

$$
\begin{aligned}
\Phi(2) &= E[(\nabla^2 z_k)(\nabla^2 z_{k-2})] \\
&= T^2 q + R
\end{aligned} \tag{10-4-32}
$$

根据各态历经性质，求出相关函数的估计

$$
\left.\begin{aligned}
\hat{\Phi}_N(0) &= \frac{1}{N} \sum_{i=k-N+1}^{k} (\nabla^2 z_i)^2 \\
\hat{\Phi}_N(1) &= \frac{1}{N} \sum_{i=k-N+1}^{k} (\nabla^2 z_k)(\nabla^2 z_{k-1}) \\
\hat{\Phi}_N(2) &= \frac{1}{N} \sum_{i=k-N+1}^{k} (\nabla^2 z_k)(\nabla^2 z_{k-2})
\end{aligned}\right\} \tag{10-4-33}
$$

解联立方程，得

$$
\left.\begin{aligned}
\hat{R} &= \frac{1}{5}[\hat{\Phi}_N(2) - \hat{\Phi}_N(1)] \\
\hat{q} &= \frac{1}{T^2}\left[\frac{4}{5}\hat{\Phi}_N(2) + \frac{1}{5}\hat{\Phi}_N(1)\right] \\
\hat{Q} &= \frac{1}{T^2}[\hat{\Phi}_N(0) + \hat{\Phi}_N(1) - 2\hat{\Phi}_N(2)]
\end{aligned}\right\} \tag{10-4-34}
$$

相应卡尔曼滤波器中的预报应改为

$$\hat{X}_{k+1/k} = \Phi \hat{X}_k + \Gamma \hat{q} \tag{10-4-35}$$

其余的$P_{k+1/k}$、K_{k+1}、$\hat{X}_{k+1}$、P_{k+1}公式不变。

三阶定常模型情况

设有三阶定常模型如下：

$$
\left.\begin{aligned}
\begin{bmatrix} x_k \\ \dot{x}_k \\ \ddot{x}_k \end{bmatrix} &= \begin{bmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{k-1} \\ \dot{x}_{k-1} \\ \ddot{x}_{k-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ w_{k-1} \end{bmatrix} \\
z_k &= x_k + v_k
\end{aligned}\right\} \tag{10-4-36}
$$

其中w_k、v_k为平稳零均值白噪声，其方差Q、R为未知，今用差分法求其估计$\hat{Q}$、$\hat{R}$。

因为
$$
\nabla^3 z_k = \nabla^3 x_k + \nabla^3 v_k \tag{10-4-37}
$$

由信号模型得

$$
\begin{aligned}
\nabla^2 x_k &= \nabla x_k - \nabla x_{k-1} = (x_k - x_{k-1}) - (x_{k-1} - x_{k-2}) \\
&= T\dot{x}_{k-1} + \frac{T^2}{2}\ddot{x}_{k-1} - T\dot{x}_{k-2} - \frac{T^2}{2}\ddot{x}_{x-2} \\
&= T^2 \dot{x}_{k-2} + \frac{T^2}{2} w_{k-2}
\end{aligned}
$$

$$
\begin{aligned}
\nabla^3 x_k &= \nabla^2 x_k - \nabla^2 x_{k-1} \\
&= T^2 \ddot{x}_{k-2} + \frac{T^2}{2} w_{k-2} - \left(T^2 \ddot{x}_{k-3} + \frac{T^2}{2} w_{k-3}\right) \\
&= \frac{T^2}{2} w_{k-2} + \frac{T^2}{2} w_{k-3}
\end{aligned} \tag{10-4-38}
$$

$$
\begin{aligned}
\nabla^2 v_k &= v_k - 2v_{k-1} + v_{k-2} \\
\nabla^3 v_k &= \nabla^2 v_k - \nabla^2 v_{k-1} \\
&= v_k - 3v_{k-1} + 3v_{k-2} - v_{k-3}
\end{aligned} \tag{10-4-39}
$$

于是有

$$
\nabla^3 z_k = \frac{T^2}{2} w_{k-2} + \frac{T^2}{2} w_{k-3} + v_k - 3v_{k-1} + 3v_{k-2} - v_{k-3} \tag{10-4-40}
$$

显见，$\nabla^3 z_k$为噪声的线性组合。当各噪声为平稳零均值白噪声时，则$\nabla^3 z_k$为平稳零均值有色噪声，其相关函数为

$$
\left.\begin{aligned}
\Phi(0) &= E[(\nabla^3 z_k)^2] = \frac{T^4}{2} Q + 20R \\
\Phi(1) &= E[(\nabla^3 z_k)(\nabla^3 z_{k-1})] = \frac{T^4}{4} Q - 15R
\end{aligned}\right\} \tag{10-4-41}
$$

再求出$\Phi(0)$和$\Phi(1)$的估计

$$
\left.\begin{aligned}
\hat{\Phi}_N(0) &= \frac{1}{N} \sum_{i=k-N+1}^{k} (\nabla^3 z_i)^2 \\
\hat{\Phi}_N(1) &= \frac{1}{N} \sum_{i=k-N+1}^{k} (\nabla^3 z_i)(\nabla^3 z_{i-1})
\end{aligned}\right\} \tag{10-4-42}
$$

于是，得

$$
\left.\begin{aligned}
\hat{R} &= \frac{1}{50}[\hat{\Phi}_N(0) - 2\hat{\Phi}_N(1)] \\
\hat{Q} &= \frac{2}{5T^4}[3\hat{\Phi}_N(0) + 4\hat{\Phi}_N(1)]
\end{aligned}\right\} \tag{10-4-43}
$$

§10-5 机动量χ^2分布检验与方差匹配

设有线性定常信号模型

$$\left.\begin{aligned} X_{k+1} &= \Phi X_k + \Gamma W_k \\ Z_{k+1} &= H X_{k+1} + V_{k+1} \end{aligned}\right\} \qquad (10\text{-}5\text{-}1)$$

其中W_k、V_k为零均值独立白噪声，分别有方差矩阵 $E[W_k \ W_k^T]=Q$，$E[V_k \ V_k^T]=R$ 假定模型为完全可观测与完全可控制的，则上述系统具有最优滤波稳态解，即当滤波步数充分大之后，有

$$P_{k+1} \longrightarrow \overline{P}$$

$$K_{k+1} \longrightarrow \overline{K}$$

$$P_{k+1/k} \longrightarrow \overline{P^-}$$

其中$\overline{P}$、$\overline{K}$、$\overline{P^-}$均为常值矩阵。三者之间存在下列关系：

$$\overline{P^-} = \Phi\overline{P}\Phi^T + \Gamma Q \Gamma^T \qquad (10\text{-}5\text{-}2)$$

$$\overline{K} = \overline{P^-}H^T\ (H\overline{P^-}H^T + R)^{-1} \qquad (10\text{-}5\text{-}3)$$

$$\overline{P} = (1-\overline{K}H)\overline{P^-} \qquad (10\text{-}5\text{-}4)$$

再来看最优滤波中的预报残差，在$k+1$时刻的预报残差为

$$\varepsilon_{k+1} = Z_{k+1} - H\hat{X}_{k+1/k} \qquad (10\text{-}5\text{-}5)$$

ε_{k+1}又称为新息序列。由第七章知道，当滤波器工作在最优情况时，ε_{k+1}为零均值平稳高斯白噪声序列，有方差矩阵

$$E[\varepsilon_{k+1}\varepsilon_{k+1}^T] = H\overline{P^-}H^T + R \qquad (10\text{-}5\text{-}6)$$

利用新息的这一重要统计性质，可对滤波器的工作情况进行检验。

χ^2分布检验法

如用$\varepsilon_{k+1}(i)$表示矢量ε_{k+1}的第i个元素，记$P_*(i)=(H\overline{P^-}H^T+R)_{i,i}$，$i=1,2,\cdots,m$ 下标(i,i)表示矩阵对角线上第i元素。

令 $$\mu_N(i) = \sum_{j=0}^{N-1}\varepsilon_{k-j}^2(i)/P_*(i) \qquad i=1,2,\cdots,m，N\text{为数据数目} \qquad (10\text{-}5\text{-}8)$$

由于$P_*(i)$就是高斯随机变量$\varepsilon_{k-j}(i)$的方差，由新息序列的统计性质可知，$\mu_N(i)$满足自由度为N的χ^2分布。

利用$\mu_N(i)$，$i=1,2,\cdots,m$，这m个统计量可以判断目标动态有否变化。作假设：目标到k时刻为止动态无变化，由χ^2分布表可查得$\mu_N(i)$大于某门限M的概率为α，α为适当选取的小概率，此时用下式作检验。

若$\mu_N(i)<M$，对$i=1,2\cdots,m$均成立时，判断原假设成立。

若$\mu_N(i)>M$，对$i=1,2,\cdots,m$至少有一个成立时，判断原假设不成立。

上述常数α就是虚警概率，门限M愈大则虚警概率越小，但漏报概率增大。所以，在选择门限时应兼顾虚警和漏报。当已判决到目标有机动后。再估计机动的强度。

方差匹配

当已判断目标动态有变化后，需要估计这种动态变化的强度。我们把此变化看作是随机干扰，此时，状态方程应加以修改，成为

$$X_{k+1}=\Phi X_k+\Gamma W_k+\Gamma_1\xi_k \tag{10-5-9}$$

式中$\Gamma_1\xi_k$为新增加的随机干扰。假定ξ_k为零均值白噪声，方差$E[\xi_k\xi_k^T]=Q'_k$，且与W_k，V_k独立。

现在要由滤波器本身来估计Q'_k。利用(10-5-6)式

$$E[\varepsilon_{k+1}\varepsilon_{k+1}^T]=HP_{k+1/k}H^T+R \tag{10-5-10}$$

其中$P_{k+1/k}$应当是实际的预报误差方差，故有

$$P_{k+1/k}=P_{k+1/k}^*+\Gamma_1 Q'_k\Gamma_1^T \tag{10-5-11}$$

其中$P_{k+1/k}^*$为未计入动态变化时计算出的预报误差方差，于是有

$$E[\varepsilon_{k+1}\varepsilon_{k+1}^T]=(HP_{k+1/k}^*H^T+R)+H\Gamma_1 Q'_k\Gamma_1^T H^T \tag{10-5-12}$$

利用时间平均代替统计平均，得

$$E[\varepsilon_{k+1}\varepsilon_{k+1}^T]=\frac{1}{N}\sum_{j=0}^{N-1}\varepsilon_{k-j+1}\varepsilon_{k-j+1}^T \tag{10-5-13}$$

从而得到Q'_k的估计为

$$H\Gamma_1\hat{Q}'_k\Gamma_1^T H^T=\frac{1}{N}\sum_{j=0}^{N-1}\varepsilon_{k-j+1}\varepsilon_{k-j+1}^T-(HP_{k+1/k}^*H^T+R) \tag{10-5-14}$$

当$(H\Gamma_1)^T(H\Gamma_1)>0$时，由上式可唯一解出

$$\begin{aligned}\hat{Q}'_k=&[(H\Gamma_1)^T(H\Gamma_1)]^{-1}(H\Gamma_1)^T\\&\times\left[\frac{1}{N}\sum_{j=0}^{N-1}\varepsilon_{k+1-j}\varepsilon_{k+1-j}^T-(HP_{k+1/k}^*H^T+R)\right]\\&\times(H\Gamma_1)[(H\Gamma_1)^T(H\Gamma_1)]^{-1}\end{aligned} \tag{10-5-15}$$

求得$\hat{Q}'_k$后，代入(10-5-11)式，即可算得实际的预报误差方差$P_{k+1/k}$，利用此$P_{k+1/k}$代入卡尔曼滤波器中就可得到修正后的状态估计。

上述方法是由线性定常模型的稳态导出的。对于线性非定常模型，新息序列$\{\varepsilon_k\}$仍是零均值高斯白噪声序列，但由于不存在稳态的$\overline{P}$、$\overline{K}$、$\overline{P}$，因此ε_k的方差随时间而变，从而$\{\varepsilon_k\}$序列是非平稳序列。但是，当滤波器中的$P_{k+1/k}$变化不快时，只要计算速度跟得上，且(10-5-13)式能够成立，则上述自适应算法仍是近似适用的。

§10-6 噪声统计量与信号状态的交替估计

在实际工程问题中应用卡尔曼滤波时，常常会遇到这样情况，不仅状态初值不能确切知道，连噪声统计量也不能确切知道，即就需要对信号状态与噪声统计量同时进行在线估计。本节中我们用次优递推线性估计方法对噪声统计量进行自适应估计。噪声统计量的估计与信号状态的估计是交替进行的，本节给出了交替估计法自适应卡尔曼滤波的递推公式。

信号状态的估计

设有定常信号模型

$$X_k = \Phi X_{k-1} + W_k \tag{10-6-1}$$

$$Z_k = H X_k + V_k \tag{10-6-2}$$

其中W_k、V_k为不相关白噪声，W_k的均值q方差Q为未知，V_k的均值r、方差R也为未知，但已知在t_{k-1}时刻的估计$\hat{q}_{k-1}$、$\hat{Q}_{k-1}$、$\hat{r}_{k-1}$、$\hat{R}_{k-1}$、$\hat{X}_{k-1}$、P_{k-1}。

在得到观测值Z_k之后，利用卡尔曼滤波器可以得到t_k时刻的状态估计$\hat{Z}_k$、P_k与增益K_k、新息ε_k。

$$\hat{X}_{k/k-1} = \Phi \hat{X}_{k-1} + \hat{q}_{k-1} \tag{10-6-3}$$

$$P_{k/k-1} = \Phi P_{k-1} \Phi^T + \hat{Q}_{k-1} \tag{10-6-4}$$

$$K_k = P_{k/k-1} H^T (H P_{k/k-1} H^T + \hat{R}_{k-1}) \tag{10-6-5}$$

$$\hat{\varepsilon}_k = Z_k - \hat{r}_{k-1} - H \hat{X}_{k/k-1} \tag{10-6-6}$$

$$\hat{X}_k = \hat{X}_{k/k-1} + K_k \hat{\varepsilon}_k \tag{10-6-7}$$

$$P_k = (1 - K_k H) P_{k/k-1} (1 - K_k H)^T + K_k \hat{R}_{k-1} R_k^T \tag{10-6-8}$$

注意这里用$\hat{r}_{k-1}$、$\hat{R}_{k-1}$代替了原来卡尔曼滤波器中的r_k与R_k。P_k的表达式采用(10-6-8)形式，是考虑到自适应过程中K_k不是最佳增益。

噪声统计量的估计

1.对q的估计

由 $$E[X_k] = \Phi E[X_{k-1}] + q$$

得 $$q = E[X_k - \Phi X_{k-1}] \tag{10-6-9}$$

用时间平均代替统计平均，并用$\hat{X}_i$、$\hat{X}_{i-1}$代替X_i、X_{i-1}，得

$$\hat{q}_k = \frac{1}{k} \sum_{i=1}^{k} \left(\hat{X}_i - \Phi \hat{X}_{i-1} \right)$$

$$= \left(1 - \frac{1}{k}\right) \hat{q}_{k-1} + \frac{1}{k} \left(\hat{X}_k - \Phi \hat{X}_{k-1} \right) \tag{10-6-10}$$

2.对Q的估计

已知 $$X_k - \hat{X}_k = X_k - \hat{X}_{k/k-1} - K_k \varepsilon_k$$

令 $$\tilde{X}_k = X_k - \hat{X}_k, \quad \tilde{X}_{k/k-1} = X_k - \hat{X}_{k/k-1}$$

得 $$\tilde{X}_{k/k-1} = \tilde{X}_k + K_k \varepsilon_k \tag{10-6-11}$$

两边取均方值，得

$$\Phi P_{k-1} \Phi^T + Q = P_k + E[K_k \varepsilon_k \varepsilon_k^T K_k^T]$$

$$+ E[K_k \varepsilon_k \tilde{X}_k^T] + E[\tilde{X}_k \varepsilon_k^T K_k^T] \tag{10-6-12}$$

其中 $$E[K_k \varepsilon_k \tilde{X}_k^T] = K_k H P_{k/k-1} (1 - K_k H)^T - K_k P K_k^T$$

$$= -(1 - K_k H) P_{k/k-1} (1 - K_k H)^T + P_{k/k-1} (1 - K_k H)^T - K$$

$$= P_{k/k-1} (1 - K_k H)^T - P_k \tag{10-6-13}$$

同理 $E[\widetilde{X}\varepsilon_k^T K_k^T] = (I - K_k H) P_{k/k-1} - P_k$ (10-6-14)

用时间平均代替统计平均，用$\hat{\varepsilon}_i$代替ε_i，得

$$\hat{Q}_k = \frac{1}{k}\sum_{i=1}^{k}\Big[K_i\hat{\varepsilon}_i\hat{\varepsilon}_i^T K_i^T - \Phi P_{i-1}\Phi^T - P_i + P_{i/i}(I - K_i H)^T + (I - K_i H)P_{i/i-1}\Big]$$

$$= \left(1 - \frac{1}{k}\right)\hat{Q}_{k-1} + \frac{1}{k}\Big[K_k\hat{\varepsilon}_k\hat{\varepsilon}_k^T K_k^T - \Phi P_{k/k-1}\Phi^T - P_k + P_{k/k-1}(I - K_k H)^T + (I - K_k H)P_{k/k-1}\Big] \tag{10-6-15}$$

3.对r的估计

由$r = E[Z_k - HX_k]$，用时间平均代替统计平均，用$\hat{X}_i$代替X_i，得

$$\hat{r}_k = \frac{1}{k}\sum_{i=1}^{k}(Z_i - H\hat{X}_i)$$

$$= \left(1 - \frac{1}{k}\right)\hat{r}_{k-1} + \frac{1}{k}\left(Z_k - H\hat{Z}_k\right) \tag{10-6-16}$$

4.对R的估计

由 $E[\varepsilon_k\varepsilon_k^T] = HP_{k/k-1}H^T + R$ 得

$$R = E[\varepsilon_k\varepsilon_k^T - HP_{k/k-1}H^T]$$

用时间平均代替统计平均，用$\hat{\varepsilon}_i$代替ε_i，得

$$\hat{R}_k = \frac{1}{k}\sum_{i=1}^{k}\left(\hat{\varepsilon}_i\hat{\varepsilon}_i^T - HP_{k/k-1}H^T\right)$$

$$= \left(1 - \frac{1}{k}\right)\hat{R}_{k-1} + \frac{1}{k}\left(\hat{\varepsilon}_k\hat{\varepsilon}_k^T - HP_{k/k-1}H^T\right) \tag{10-6-18}$$

(10-6-10)、(10-6-15)、(10-6-16)、(10-6-18) 四式构成了噪声统计量q、Q、r、R的估计算法，不难证明它们都是无偏估计。

注意到当滤波趋近稳态时，有$P_k = (I - K_k H)P_{k/k-1}$，故(10-6-15)式可近似化简为

$$\hat{Q}_k \approx \left(1 - \frac{1}{k}\right)\hat{Q}_{k-1} + \frac{1}{k}\left(K_k\hat{\varepsilon}_k\hat{\varepsilon}_k^T K_k^T - \Phi P_{k/k-1}\Phi^T + P_k\right) \tag{10-6-19}$$

上述噪声估计器是针对 q、Q、r、R为常量情况导出的，当它们作缓慢变化时，可以应用渐消记忆滤波原理，用阻尼因子b_k代替系数$\frac{1}{k}$，这里取$b_k = \frac{1-a}{1-a^{k+1}}$，$0<0.95<a<1$。

最后总结一下交替估计法自适应卡尔曼滤波的算法与流程，有

$$\hat{X}_{k/k-1} = \Phi\hat{X}_{k-1} + \hat{q}_{k-1} \tag{10-6-20}$$

$$P_{k/k-1} = \Phi P_{k-1}\Phi^T + \hat{Q}_{k-1} \tag{10-6-21}$$

$$K_k = P_{k/k-1}H^T(HP_{k/k-1}H^T + \hat{R}_{k-1})^{-1} \tag{10-6-22}$$

$$\hat{\varepsilon}_k = Z_k - \hat{r}_{k-1} - H\hat{X}_{k/k-1} \tag{10-6-23}$$

$$\hat{X}_k = \hat{X}_{k/k-1} + K_k\hat{\varepsilon}_k \tag{10-6-24}$$

$$P_k = (I - K_kH)P_{k/k-1}(I - K_kH)^T + K_k\hat{R}_{k-1}K_k^T \tag{10-6-25}$$

$$b_k = \frac{1-\alpha}{1-\alpha^{k+1}}, \qquad 0<\alpha<1 \tag{10-6-26}$$

$$\hat{q}_k = \hat{q}_{k-1} + b_k(\hat{X}_k - \Phi\hat{X}_{k-1} - \hat{q}_{k-1}) \tag{10-6-27}$$

$$\hat{Q}_k = \hat{Q}_{k-1} + b_k(P_k + K_k\hat{\varepsilon}_k\hat{\varepsilon}_k^TK_k^T - \Phi P_{k-1}\Phi^T - \hat{Q}_{k-1}) \tag{10-6-28}$$

$$\hat{r}_k = \hat{r}_{k-1} + b_k(Z_k - H\hat{X}_k - \hat{r}_{k-1}) \tag{10-6-29}$$

$$\hat{R}_k = \hat{R}_{k-1} + b_k(\hat{\varepsilon}_k\hat{\varepsilon}_k^T - HP_{k/k-1}H^T - \hat{R}_{k-1}) \tag{10-6-30}$$

初始值　$\hat{X}_0 = E[X_0]$，$P_0 = \mathrm{var}[X_0]$，$\hat{q}_0 = E[q_0]$，$\hat{Q}_0 = E[Q_0]$，$\hat{r}_0 = E[r_0]$，$\hat{R}_0 = E[R_0]$。

加权系数b_k也可以取成其它形式，例如，令$b_k = \alpha^{k-1} - \alpha + 1$，$0<\alpha<1$。也可取得良好效果。

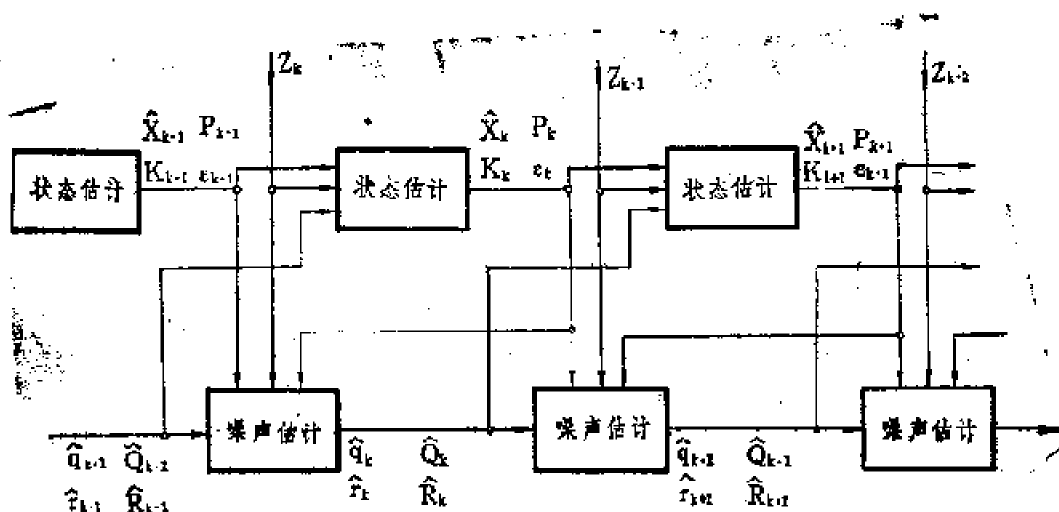


图10-6-1　交替估计法自适应卡尔曼滤波流程图

应当指出，以上算法是针对定常信号模型（噪声统计为常量）导出的，但由于算法是递推形式的，当噪声统计作缓慢变化时，只要加权系数取得合适，滤波器仍能实时跟踪噪声统计的变化，保持滤波器的收敛。

还应当指出，这种算法的稳定性尚未得到证明，初值应当仔细选取，以保证滤波的收敛。

虚拟噪声原理

在许多实际情况下，不仅对噪声统计缺乏先验知识，而且对Φ、H等模型参数也不确切知道。

假定给出的信号模型为

$$\left.\begin{aligned} X_{k+1} &= \Phi X_k + W_k \\ Z_{k+1} &= H X_{k+1} + V_{k+1} \end{aligned}\right\} \tag{10-6-31}$$

而真实的信号模型应为

$$\left.\begin{aligned} X_{k+1} &= (\Phi + \Delta\Phi) X_k + W_k \\ Z_{k+1} &= (H + \Delta H) X_{k+1} + V_{k+1} \end{aligned}\right\} \tag{10-6-32}$$

其中$\Delta\Phi$、ΔH表示模型误差部分，为了补偿模型误差，令

$$\left.\begin{aligned} X_{k+1} &= \Phi X_k + U_k \\ Z_{k+1} &= H X_{k+1} + S_{k+1} \\ \text{其中}\quad U_k &= \Delta\Phi X_k + W_k,\ S_{k+1} = \Delta H X_{k+1} + V_{k+1} \end{aligned}\right\} \tag{10-6-33}$$

我们把U_k、S_k称为虚拟噪声。然后把含有模型误差的模型转化成含有虚拟噪声的模型。由于U_k、S_k中含有模型误差，模型误差与信号状态有关。因此，虚拟噪声是时变的未知均值和方差的噪声，并且严格说来是非白的。当$\Delta\phi$、ΔA、相对很小时，可近似地把U_k、S_k看成白噪声，于是可以应用交替估计法求得虚拟噪声和状态的估计。

$$\left.\begin{aligned} \hat{X}_{k+1/k} &= \Phi \hat{X}_k + \hat{q}_k \\ P_{k+1/k} &= \Phi P_k \Phi^T + \hat{Q}_k \\ K_{k+1} &= P_{k+1/k} H^T (H P_{k+1/k} H^T + \hat{R}_k)^{-1} \\ \hat{X}_{k+1} &= \hat{X}_{k+1/k} + K_{k+1} \varepsilon_{k+1} \\ \varepsilon_{k+1} &= Z_{k+1} - H \hat{X}_{k+1/k} - \hat{r}_k \\ P_{k+1} &= (1 - K_{k+1} H) P_{k+1/k} \end{aligned}\right\} \tag{10-6-34}$$

注意，已用$\hat{r}_k$、$\hat{R}_k$代替r_{k+1}、R_{k+1}

$$\left.\begin{aligned} \hat{q}_{k+1} &= (1 - b_k)\hat{q}_k + b_k [\hat{X}_{k+1} - \Phi \hat{X}_k] \\ \hat{Q}_{k+1} &= (1 - b_k)\hat{Q}_k + b_k [K_{k+1} \varepsilon_{k+1} \varepsilon_{k+1}^T K_{k+1}^T \\ &\quad + P_{k+1} - \Phi P_k \Phi^T] \\ \hat{r}_{k+1} &= (1 - b_k)\hat{r}_k + b_k [Z_{k+1} - H \hat{X}_{k+1}] \\ \hat{R}_{k+1} &= (1 - b_k)\hat{R}_k + b_k [\varepsilon_{k+1} \varepsilon_{k+1}^T - H P_{k+1/k} H^T] \\ b_k &= \frac{1 - a}{1 - a^{k+1}} \qquad 0 < a < 1 \end{aligned}\right\} \tag{10-6-35}$$

初始条件为$\hat{X}_0$、P_0、$\hat{q}_0$、$\hat{Q}_0$、$\hat{r}_0$、$\hat{R}_0$

其中 $E[U_k] = \hat{q}_k$，　$E[U_k U_k^T] = \hat{Q}_k$

$E[S_k] = \hat{r}_k$，　$E[S_k S_k^T] = \hat{R}_k$

§10-7 信号状态与模型参数的联合估计

在例10-3-1中研究了一个信号状态与模型参数联合估计的简单例子，本节中研究普遍性理论。在例10-3-1中，我们是采用增广状态变量与广义卡尔曼滤波方法去求得状态 X_k 与参数 a 的估计的，但是这个方法是有缺点的，一是当状态维数较高、模型中未知参数较多时计算量很大，二是广义卡尔曼滤波法精度较低（忽略了线性化余项），容易造成滤波发散。这里采用上节中虚拟噪声原理来解决信号状态与模型参数的联合估计问题。

设有线性离散模型

$$X_{k+1}=\Phi(\theta)X_k+B(\theta)U_k+W_k \tag{10-7-1}$$

$$Z_k=H_kX_k+V_k \tag{10-7-2}$$

在 Φ、B 中含有未知模型参数矢量 θ，W_k 与 V_k 为互相独立的白色序列；但它们的均值和方差均为未知。

如果参数 θ 是时不变的，其动态模型为

$$\theta_{k+1}=\theta_k$$

如果参数 θ 是时变的，其动态模型为

$$\theta_{k+1}=F\theta_k+\xi_k$$

噪声 ξ_k 为白色序列，与 W_k、V_k 不相关，其均值和方差为未知。

现在要根据直到 k 时刻的观测序列 Z_1，Z_2，Z_3，…，Z_k 对信号状态 Z_k 与模型参数 θ_k 作出估计。

令
$$E[W_k]=q_k,\qquad E[W_kW_k^T]=Q_k$$
$$E[V_k]=r_k,\qquad E[V_kV_k^T]=R_k$$
$$E[\xi_k]=s_k,\qquad E[\xi_k\xi_k^T]=S_k$$

把(10-7-1)式代入(10-7-2)式，得

$$Z_k=H_k\Phi(\theta)X_{k-1}+H_kB(\theta)U_{k-1}+H_kW_{k-1}+V_k \tag{10-7-3}$$

通常 Φ、B 中不是所有元素为未知，而是部分未知部分已知，因此可以把 $H_k\Phi(\theta)X_{k-1}$ 与 $H_kB(\theta)U_{k-1}$ 加以分解，把(10-7-3)式改写成

$$\begin{aligned}Z_k=&C(X_{k-1},\ U_{k-1})\theta+D(X_{k-1},\ U_{k-1})\\&+H_kW_{k-1}+V_k\end{aligned} \tag{10-7-4}$$

在(10-3-4)中以滤波估计 $\hat{X}_{k-1}$ 近似代替 X_{k-1}，则可得到关于 θ 的观测方程

$$Z_k=C(\hat{X}_{k-1},\ U_{k-1})\theta_k+D(\hat{X}_{k-1},\ U_{k-1})+\eta_k \tag{10-7-5}$$

其中　$\eta_k=H_kW_{k-1}+V_k+$ 模型误差，

这里模型误差是由于近似代替引入的，把 η_k 看成是虚拟噪声，显然 η_k 具有未知的均值与方差，即使 W_k、V_k 是平稳的，η_k 也是非平稳的。

令
$$E[\eta_k]=n_k,\quad E[\eta_k\eta_k^T]=N_k$$

于是得到模型参数 θ 的信号模型为

$$\theta_{k+1}=\theta_k+\xi_k \tag{10-7-6}$$

与

$$Z_k=C(\hat{X}_{k-1},\ U_{k-1})\theta_k+D(\hat{X}_{k-1},\ U_{k-1})+\eta_k \tag{10-7-7}$$

根据此模型，构造估计θ的卡尔曼滤波器如下：

$$\hat{\theta}_{k+1} = \hat{\theta}_{k+1/k} + K_{\theta(k+1)} \cdot \varepsilon_{\theta(k+1)} \tag{10-7-8}$$

$$\hat{\theta}_{k+1/k} = F\hat{\theta}_k + \hat{s}_k \tag{10-7-9}$$

$$\varepsilon_{\theta(k+1)} = Z_{k+1} - C_k\hat{\theta}_{k+1/k} - D_k - \hat{n}_k \tag{10-7-10}$$

$$P_{\theta(k+1/k)} = FP_{\theta(k)}F^T + \hat{S}_k \tag{10-7-11}$$

$$K_{\theta(k+1)} = P_{\theta(k+1/k)}C_k^T[C_kP_{\theta(k+1/k)}C_k^T + N_k^T]^{-1} \tag{10-7-12}$$

$$P_{\theta(k+1)} = (I - K_{\theta(k+1)}C_k)P_{\theta(k+1/k)} \tag{10-7-13}$$

其中虚拟噪声的估计器为

$$\hat{s}_{k+1} = (1-\beta_k)\hat{s}_k + \beta_k[\hat{\theta}_{k+1} - F\hat{\theta}_k] \tag{10-7-14}$$

$$\hat{S}_{k+1} = (1-\beta_k)\hat{S}_k + \beta_k[K_{\theta(k+1)}\varepsilon_{\theta(k+1)}\varepsilon^T_{\theta(k+1)}K^T_{\theta(k+1)} + P_{\theta(k+1)} - FP_{\theta(k)}F^T] \tag{10-7-15}$$

$$\hat{n}_{k+1} = (1-\beta_k)\hat{n}_k + \beta_k[Z_{k+1} - C_k\hat{\theta}_{k+1/k} - D_k] \tag{10-7-16}$$

$$\hat{N}_{k+1} = (1-\beta_k)\hat{N}_k + \beta_k[\varepsilon_{\theta(k+1)}\varepsilon^T_{\theta(k+1)} - C_kP_{\theta(k+1/k)}C_k^T] \tag{10-7-17}$$

$$C_k = C(\hat{X}_k,\ U_k),\quad D_k = D(\hat{X}_k,\ U_k) \tag{10-7-18}$$

上文的意思是说如果已知k时刻上信号状态估计就可以得出$k+1$时刻的模型参数θ的估计与参数噪声ξ_k的估计及虚拟噪声η_k的估计。在上述算法中，η_k近似看作白噪声。

然后在(10-7-1)式中，θ用前面得到的$\hat{\theta}_{k+1}$近似代替，可得

$$X_{k+1} = \Phi(\hat{\theta}_{k+1})X_k + B(\hat{\theta}_{k+1})U_k + \lambda_k$$

其中$\lambda_k = W_k +$模型误差，模型误差由近似代替所引起。令

$$E[\lambda_k] = g_k,\quad E[\lambda_k\lambda_k^T] = G_k$$

显然λ_k是具有未知均值与方差的噪声，近似地看作是白噪声，于是得到信号模型

$$X_{k+1} = \Phi(\hat{\theta}_{k+1})X_k + B(\hat{\theta}_{k+1})U_k + \lambda_k \tag{10-7-19}$$

$$Z_{k+1} = H_{k+1}X_{k+1} + V_{k+1} \tag{10-7-20}$$

根据上述模型构造信号X_k的估计器如下：

$$\hat{X}_{k+1} = \hat{X}_{k+1/k} + K_{x(k+1)}\varepsilon_{x(k+1)} \tag{10-7-21}$$

$$\hat{X}_{k+1/k} = \Phi(\hat{\theta}_{k+1})\hat{X}_k + B(\hat{\theta}_{k+1})U_k + \hat{g}_k \tag{10-7-22}$$

$$\varepsilon_{x(k+1)} = Z_{k+1} - H_{k+1}\hat{X}_{k+1/k} - \hat{r}_k \tag{10-7-23}$$

$$K_{x(k+1)} = P_{x(k+1/k)}H^T_{k+1}[H_{k+1}P_{x(k+1/k)}H^T_{k+1} + \hat{R}_k]^{-1} \tag{10-7-24}$$

$$P_{x(k+1/k)} = \Phi(\hat{\theta}_{k+1})P_{x(k)}\Phi^T(\hat{\theta}_{k+1}) + \hat{Q}_k \tag{10-7-25}$$

$$P_{x(k+1)} = [I - K_{x(k+1)}H_{k+1}]P_{x(k+1/k)} \tag{10-7-26}$$

测量噪声与虚拟噪声λ_k的统计量估计为

$$\hat{g}_{k+1} = (1-\beta_k)\hat{g}_k + \beta_k[\hat{X}_{k+1} - \Phi(\hat{\theta}_{k+1})\hat{X}_k - B(\hat{\theta}_{k+1})U_k] \tag{10-7-27}$$

$$\hat{G}_{k+1}=(1-\beta_k)\hat{G}_k+\beta_k[K_{x(k+1)}\varepsilon_{x(k+1)}\varepsilon_x^T{}_{(k+1)}K_x^T{}_{(k+1)}$$

$$+P_{x(k+1)}-\Phi(\theta_{k+1})P_{x(k)}\Phi^T(\theta_{k+1})] \tag{10-7-28}$$

$$\hat{r}_{k+1}=(1-\beta_k)\hat{r}_k+\beta_k[Z_{k+1}-H_{k+1}\hat{X}_{k+1/k}] \tag{10-7-29}$$

$$\hat{R}_{k+1}=(1-\beta_k)\hat{R}_k+\beta_k[\varepsilon_{x(k+1)}\varepsilon_x^T{}_{(k+1)}$$

$$-H_{k+1}P_{x(k+1/k)}H_{k+1}^T] \tag{10-7-30}$$

其中β_k可取

$$\beta_k=\alpha^{1+k}+1-\alpha, \qquad 0<\alpha<1$$

或

$$\beta_k=\frac{1-\alpha}{1-\alpha^{k+1}}, \qquad 0<\alpha<1$$

给出初值$\hat{X}_0$、$P_{x(0)}$、$\hat{\theta}_0$、$P_{\theta(0)}$、$\hat{s}_0$、$\hat{S}_0$、$\hat{n}_0$、$\hat{N}_0$、$\hat{g}_0$、$\hat{G}_0$、$\hat{r}_0$、$\hat{R}_0$之后，就可以交替地得到模型参数θ_k与信号状态X_k的估计。下面举一个例子说明上述自适应滤波器的应用。

航行体的自适应跟踪

用雷达跟踪飞机，用声纳跟踪船舶等航行体时，一般不能确切地知道初值与机动情况，需要采用自适应跟踪算法，飞机或船舶一类的航行体在一个空间坐标上的信号模型可表示为

$$\left.\begin{aligned}\begin{pmatrix}x_k\\ \dot{x}_k\\ \ddot{x}_k\end{pmatrix}&=\begin{pmatrix}1 & T & \frac{T^2}{2}\\ 0 & 1 & T\\ 0 & 0 & \rho\end{pmatrix}\begin{pmatrix}x_{k-1}\\ \dot{x}_{k-1}\\ \ddot{x}_{k-1}\end{pmatrix}+\begin{pmatrix}0\\0\\1\end{pmatrix}w_{k-1}\\ z_k&=[1,\ 0,\ 0]\begin{pmatrix}x_k\\ \dot{x}_k\\ \ddot{x}_k\end{pmatrix}+v_k\end{aligned}\right\} \tag{10-7-31}$$

其中噪声的统计量q_k、Q_k、r_k、R_k均为未知量。

已知

$$\Phi=\begin{pmatrix}1 & T & \frac{T^2}{2}\\ 0 & 1 & T\\ 0 & 0 & \rho\end{pmatrix} \qquad \Gamma=\begin{pmatrix}0\\0\\1\end{pmatrix}$$

$$H=[1,0,0],\ 0<\rho<1\ (\text{假设}\rho\text{值已知})$$

应用上述自适应算法就可以求得$\hat{q}_k$、$\hat{Q}_k$、$\hat{r}_k$、$\hat{R}_k$和$\hat{X}_k$、P_k的各步递推估计。

实际上航行体是在平面或空间内运动的，例如对于船舶可以看成是在平面运动的，可设置状态方程为

$$\begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \\ y \\ \dot{y} \\ \ddot{y} \end{pmatrix}_{k+1} = \begin{pmatrix} 1 & T & \frac{T^2}{2} & & & \\ 0 & 1 & T & & 0 & \\ 0 & 0 & \rho_x & & & \\ & & & 1 & T & \frac{T^2}{2} \\ & 0 & & 0 & 1 & T \\ & & & 0 & 0 & \rho_y \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \\ y \\ \dot{y} \\ \ddot{y} \end{pmatrix}_k + \begin{pmatrix} 0 \\ 0 \\ w_x \\ 0 \\ 0 \\ w_y \end{pmatrix}_k \qquad (10\text{-}7\text{-}32)$$

测量通常是在极坐标内进行的，测量值是斜距r_k和方位角β_k，得出测量方程为

$$\begin{pmatrix} r\sin\beta \\ r\cos\beta \end{pmatrix}_{k+1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \\ y \\ \dot{y} \\ \ddot{y} \end{pmatrix}_{k+1} + \begin{pmatrix} v_x \\ v_y \end{pmatrix}_{k+1} \qquad (10\text{-}7\text{-}33)$$

其中噪声w_x、w_y、v_x、v_y具有未知的均值和方差，这些均值和方差通常作缓慢时变。

为了减少计算量，可将上述模型解耦，分解成x方向和y方向两个模型。

$$\left.\begin{aligned} \begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \end{pmatrix}_{k+1} &= \begin{pmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & \rho_x \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \end{pmatrix}_k + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} w_{xk} \\ r_{k+1}\sin\beta_{k+1} &= [1,\ 0,\ 0] \begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \end{pmatrix}_{k+1} + v_{x(k+1)} \end{aligned}\right\} \qquad (10\text{-}7\text{-}34)$$

和

$$\left.\begin{aligned} \begin{pmatrix} y \\ \dot{y} \\ \ddot{y} \end{pmatrix}_{k+1} &= \begin{pmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & \rho_y \end{pmatrix} \begin{pmatrix} y \\ \dot{y} \\ \ddot{y} \end{pmatrix}_k + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} w_{yk} \\ r_{k+1}\cos\beta_{k+1} &= [1,\ 0,\ 0] \begin{pmatrix} y \\ \dot{y} \\ \ddot{y} \end{pmatrix}_{k+1} + v_{y(k+1)} \end{aligned}\right\} \qquad (10\text{-}7\text{-}35)$$

然后分别应用上述交替估计自适应算法求出估计$\hat{x}_k$、$\hat{\dot{x}}_k$、$\hat{y}_k$、$\hat{\dot{y}}_k$，再换算到极坐标系中，得

$$
\left.\begin{aligned}
\hat{r}_k &= (\hat{x}_k^2 + \hat{y}_k^2)^{\frac{1}{2}} \\
\hat{\dot{r}}_k &= (\hat{\dot{x}}_k^2 + \hat{\dot{y}}_k^2)^{\frac{1}{2}} \\
\hat{\beta}_k &= \tan^{-1}\frac{\hat{x}_k}{\hat{y}_k} \\
\hat{\dot{\beta}}_k &= \tan^{-1}\frac{\hat{\dot{x}}_k}{\hat{\dot{y}}_k}
\end{aligned}\right\} \qquad (10\text{-}7\text{-}36)
$$

加速度相关系数ρ的自适应估计

若模型中不仅噪声统计量未知，而且还有其它参数未知。航行中的飞机，船舶的加速度通常服从下述规律

$$
\ddot{x}_{k+1} = \rho \ddot{x}_k + w_k, \quad 0<\rho<1 \qquad (10\text{-}7\text{-}37)
$$

于是信号模型为

$$
\left.\begin{aligned}
\begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \end{pmatrix}_{k+1} &= \begin{pmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & \rho \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \\ \ddot{x} \end{pmatrix}_k + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} w_k \\
z_{k+1} &= x_{k+1} + v_{k+1}
\end{aligned}\right\} \qquad (10\text{-}7\text{-}38)
$$

其中ρ为未知加速度相关系数，又设w_k的均值和方差未知，v_k为零均值白噪声，但方差未知。因此，需要求出的未知量有$\hat{\rho}_k$、$\hat{q}_k$、$\hat{Q}_k$、$\hat{R}_k$、$\hat{z}_k$、P_k。

信号模型的矢量形式为

$$
\left.\begin{aligned}
\boldsymbol{X}_k &= \Phi \boldsymbol{X}_{k-1} + \Gamma \boldsymbol{W}_{k-1} \\
\boldsymbol{Z}_k &= \boldsymbol{H}\boldsymbol{X}_k + \boldsymbol{V}_k
\end{aligned}\right\} \qquad (10\text{-}7\text{-}39)
$$

对状态方程式取Z变换，得

$$
\boldsymbol{X}(Z) = [1 - Z\Phi^{-1}]^{-1}\Gamma Z^{-1}\boldsymbol{W}(Z) \qquad (10\text{-}7\text{-}40)
$$

而 $[1-Z^{-1}\Phi]^{-1} = \dfrac{1}{(1-\rho Z^{-1})(1-Z^{-1})^2}$

$$
\times \begin{pmatrix} (1-\rho Z^{-1})(1-Z^{-1}) & TZ^{-1}(1-\rho Z^{-1}) & \frac{T^2}{2}Z^{-1}(1+Z^{-1}) \\ 0 & (1-\rho Z^{-1})(1-Z^{-1}) & TZ^{-1}(1-Z^{-1}) \\ 0 & 0 & (1-Z^{-1})^2 \end{pmatrix} \qquad (10\text{-}7\text{-}41)
$$

把Z^{-1}看作是移位算子，得

$$
\begin{aligned}
&(1-\rho Z^{-1})(1-Z^{-1})^2 \boldsymbol{Z}_k \\
&= \frac{T^2}{2}Z^{-1}(1+Z^{-1})\boldsymbol{W}_{k+1} + (1-\rho Z^{-1})(1-Z^{-1})^2\boldsymbol{V}_k
\end{aligned} \qquad (10\text{-}7\text{-}42)
$$

记新测量值y_k

$$
y_k = (1-Z^{-1})^2\boldsymbol{Z}_k = \boldsymbol{Z}_k - 2\boldsymbol{Z}_{k-1} + \boldsymbol{Z}_{k-2} \qquad (10\text{-}7\text{-}43)
$$

并令噪声项为η_k，则有

$$\eta_k=\frac{T^2}{2}Z^{-1}(1+Z^{-1})W_{k-1}+(1-\rho Z^{-1})(1-Z^{-1})^2V_k$$

$$=\frac{T^2}{2}\quad(W_{k-2}+W_{k-3})$$

$$+V_k-(2+\rho)V_{k-1}+(1+2\rho)V_{k-2}-\rho V_{k-3} \tag{10-7-44}$$

η_k是两个滑动平均(MA)过程之和，可由一个等价的滑动平均过示表示。令

$$\eta_k=\xi_k+a_1\xi_{k-1}+a_2\xi_{k-2}+a_3\xi_{k-3} \tag{10-7-45}$$

ξ_k为零均值，方差未知的白噪声，a_1、a_2、a_3为未知系数。

由(10-7-42)式得

$$y_k-\rho y_{k-1}=\eta_k$$

$$=\xi_k+a_1\xi_{k-1}+a_2\xi_{k-2}+a_3\xi_{k-3} \tag{10-7-46}$$

于是，得到以y_k为变量的ARMA模型，y_k称为辅助变量，注意这里W_k、V_k、η_k、ξ_k、y_k、Z_k，均是标量。

记参数矢量

$$\theta_k=\begin{pmatrix}\rho_k\\ a_{1k}\\ a_{2k}\\ a_{3k}\end{pmatrix}\qquad h_k=[y_{k-1},\ \xi_{k-1},\ \xi_{k-2},\ \xi_{k-3}] \tag{10-7-47}$$

则(10-7-46)式可改写成

$$y_k=h_k^T\theta_k+\xi_k \tag{10-7-48}$$

利用第二章中递推加权最小二乘估计公式，得

$$\hat{\theta}_k=\hat{\theta}_{k-1}+M_k[Y_k-h_k\hat{\theta}_{k-1}] \tag{10-7-49}$$

$$M_k=L_{k-1}h_k[1-h_kL_{k-1}h_k^T]^{-1} \tag{10-7-50}$$

$$L_k=(1-M_kh_k)L_{k-1} \tag{10-7-51}$$

初值取$\hat{\theta}_0$、L_0，通常可取$\hat{\theta}_0=0$，L_0=取大的正数。

一旦得到参数ρ_k的估计$\hat{\rho}_k$后，把它代入(10-6-1)式中，一般情况下ρ_k可用经验估计值代入通常取$\rho=0.8\sim0.9$

$$\begin{pmatrix}x_k\\ \dot{x}_k\\ \ddot{x}_k\end{pmatrix}=\begin{pmatrix}1 & T & \frac{T^2}{2}\\ 0 & 1 & T\\ 0 & 0 & \hat{\rho}_{k-1}\end{pmatrix}\begin{pmatrix}x_{k-1}\\ \dot{x}_{k-1}\\ \ddot{x}_{k-1}\end{pmatrix}+\begin{pmatrix}0\\ 0\\ 1\end{pmatrix}W^*_{k-1} \tag{10-7-52}$$

其中　$W^*_{k-1}=W_{k-1}$+模型误差，模型误差项包括ρ_{k-1}的估计误差引起的误差，我们把W^*_k称为虚拟噪声，它可能带有未知时变统计值。令

$$E[W^*_k]=q_k,\quad E[W^{*2}_k]=Q_k \tag{10-7-53}$$

这里q_k和Q_k均为未知。

对于(10-7-52)的模型，有如下的卡尔曼滤波方程

$$
\left.\begin{aligned}
&\text{预报} && \hat{X}_{k+1/k}=\Phi\hat{X}_k+\Gamma\hat{q}_k \\
&\text{预报方差} && P_{k+1/k}=\Phi P_k\Phi^T+\Gamma\hat{Q}_k\Gamma^T \\
&\text{增益} && K_{k+1}=P_{k+1/k}H^T[HP_{k+1/k}H^T+\hat{R}_k]^{-1} \\
&\text{滤波} && \hat{X}_{k+1}=\hat{X}_{k+1/k}+K_{k+1}[Z_{k+1}-H\hat{X}_{k+1/k}] \\
&\text{滤波方差} && P_{k+1}=(1-K_{k+1}H)P_{k+1/k}
\end{aligned}\right\} \qquad (10\text{-}7\text{-}54)
$$

其中虚拟噪声的均值$\hat{q}_k$和方差$\hat{Q}_k$，测量噪声方差$\hat{R}_k$用下述公式进行递推估计

$$
\left.\begin{aligned}
&\hat{q}_{k+1}=(1-b_{k+1})\hat{q}_k+b_{k+1}(\hat{\ddot{x}}_{k+1/k}-\hat{\rho}_k\hat{\ddot{x}}_k) \\
&\hat{Q}_{k+1}=(1-b_{k+1})\hat{Q}_k+b_{k+1}[k^2_{3(k+1)}\varepsilon^2_{k+1}+P_{33(k+1)}-\hat{\rho}_k^2P_{33(k)}] \\
&\hat{R}_{k+1}=(1-b_{k+1})\hat{R}_k+b_{k+1}[\varepsilon^2_{k+1}-P_{33(k+1/k)}]
\end{aligned}\right\} \qquad (10\text{-}7\text{-}55)
$$

其中 $\varepsilon_{k+1}=Z_{k+1}-H\hat{X}_{k+1/k}$

$K_{3(k+1)}$为K_{k+1}的第3个元素

$P_{ij(k+1)}$为P_{k+1}的第ij元素 (10-7-56)

$b_k=\dfrac{1-\alpha}{1-\alpha^{k+1}}$，$0<\alpha<1$

α称为遗忘因子，随着时间增大，b_k逐渐变小。

§10-8 数值发散及其抑制方法

从理论上讲，协方差矩阵应当是正定的和对称的，但由于计算过程中的截尾效应，可能使它失去对称性和正定性，其特征值可能出现负值，负的特征值意味着不稳定因素，最后导致滤波器发散，这种现象称为数值发散。

数值计算中截尾的影响可用下例说明之。

设有数值$e\ll1$，用记号$\underline{\underline{t}}$表示截尾运算。

$$1+e\;\underline{\underline{t}}\neq1$$

而

$$1+e^2\;\underline{\underline{t}}=1$$

例10-8-1 现有$P_{k/k-1}=\begin{pmatrix}1&0\\0&1\end{pmatrix}$

$$H_k=[1,\ 0],\quad R_k=e^2$$

则

$$P_k^{-1}=P^{-1}_{k/k-1}+H^TR_k^{-1}H$$

$$=\begin{pmatrix}1&0\\0&1\end{pmatrix}+\begin{pmatrix}\frac{1}{e^2}&0\\0&1\end{pmatrix}=\begin{pmatrix}1+\frac{1}{e^2}&0\\0&1\end{pmatrix}$$

或
$$P_k=\begin{pmatrix}\dfrac{\sigma^2}{1+\sigma^2} & 0\\ 0 & 1\end{pmatrix}$$

由于数值计算中的截尾，使得

$$P_k \doteq \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}$$

相应的增益为

$$K_k=P_kH_k^TR_k^{-1}=0$$

增益为零，意味着滤波器已经饱和。

为了克服滤波器发散，计算时可采用双倍字长，当不允许采用双倍字长时，可采用特殊的算法加以补救。

矩阵平方根

由矩阵代数知道，一个$n\times n$对称正半定矩阵，至少存在一个平方根矩阵，满足如下等式

$$\sqrt{A}\cdot(\sqrt{A})^T=A \tag{10-8-1}$$

实际上有许多个$\sqrt{A}$满足上式，平方根算法的基本思想是利用$\sqrt{A}$的递推算法代替P的递推算法，而$\sqrt{P}$的字长约为P字长的一半。

考察一个简单情况，比如$P=\sigma^2$，则$\sqrt{P}=\sigma$，显然动态范围压缩了一倍，从而避免了截尾效应，把这个基本思想推广到矢量情况。记

$$\left.\begin{aligned}&P_{k/k-1}\triangleq S_{k/k-1}S_{k/k-1}^T\\&P_k\triangleq S_kS_k^T\\&Q_k\triangleq W_kW_k^T\\&R_k\triangleq V_kV_k^T\end{aligned}\right\} \tag{10-8-2}$$

一个正半定矩阵的平方根不是唯一的，一种特别有用的形式是上三角阵或下三角阵。

把一个正半定矩阵分解成下三角阵与其转置阵之积，这种分解称为Cholesky分解。对于矩阵A，记分解后的下三角阵为$c\sqrt{A}$，则有

$$c\sqrt{A}\,(c\sqrt{A})^T=A \tag{10-8-3}$$

用A的元素A_{ij}表示，则有

$$\begin{pmatrix}c\sqrt{A_{11}} & 0 & \cdots & 0\\ c\sqrt{A_{21}} & c\sqrt{A_{22}} & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots\\ c\sqrt{A_{n1}} & c\sqrt{A_{n2}} & \cdots & c\sqrt{A_{nn}}\end{pmatrix}\begin{pmatrix}c\sqrt{A_{11}} & c\sqrt{A_{21}} & \cdots & c\sqrt{A_{n1}}\\ 0 & c\sqrt{A_{22}} & \cdots & c\sqrt{A_{n2}}\\ \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & c\sqrt{A_{nn}}\end{pmatrix}$$
$$=\begin{pmatrix}A_{11} & A_{12}\cdots & A_{1n}\\ A_{21} & A_{22}\cdots & A_{2n}\\ \cdots & \cdots & \cdots\\ A_{n1} & A_{n2}\cdots & A_{nn}\end{pmatrix} \tag{10-8-4}$$

$c\sqrt{A}$中的各元素可以逐行递推产生，令$i=1, 2, \cdots, n$，则有

$$c\sqrt{A}_{ij}=\begin{cases}\dfrac{1}{c\sqrt{A_{jj}}}\left[A_{ij}-\sum\limits_{k=1}^{j-1}c\sqrt{A_{ik}}c\sqrt{A_{jk}}\right], & j=1,2,\cdots,i-1\\ \left[A_{ii}-\sum\limits_{k=1}^{i-1}(c\sqrt{A_{ik}})^2\right]^{\frac{1}{2}} & j=i\\ 0 & j>i\end{cases} \qquad (10\text{-}8\text{-}5)$$

对A矩阵分解并产生$c\sqrt{A}$各元素的顺序如图10-8-1所示。

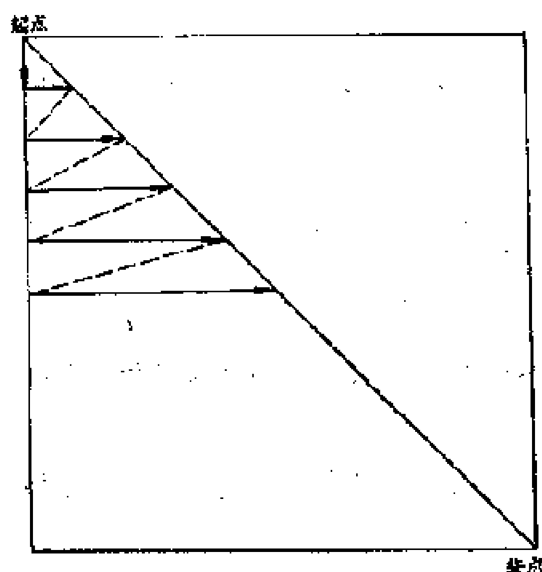


图10-8-1 对A扫描并产生$C\sqrt{A}$

例10-8-2 已知$A=\begin{pmatrix}1 & 2 & 3\\ 2 & 8 & 2\\ 3 & 2 & 14\end{pmatrix}$，求$A$的下三角平方根$c\sqrt{A}$。

解：按照(10-8-5)式递推地算出$c\sqrt{A}$的各元素为

$$c\sqrt{A_{11}}=1,\quad c\sqrt{A_{12}}=0,\quad c\sqrt{A_{13}}=0$$

$$c\sqrt{A_{21}}=\frac{1}{1}(2-0)=2$$

$$c\sqrt{A_{22}}=\sqrt{8-2^2}=2,\quad c\sqrt{A_{23}}=0$$

$$c\sqrt{A_{31}}=\frac{1}{1}(3-0)=3$$

$$c\sqrt{A_{32}}=\frac{1}{3}(2-3\times2)=-2$$

$$c\sqrt{A_{33}}=[14-3^2-(-2)^2]^{\frac{1}{2}}=1$$

最后得到

$$c\sqrt{A}=\begin{pmatrix}1 & 0 & 0\\ 2 & 2 & 0\\ 3 & -2 & 1\end{pmatrix}$$

容易验证 $c\sqrt{A}(c\sqrt{A})^T=A$。

矩阵A也可分解成上三角阵与其转置阵之积，记上三角阵为$c'\sqrt{A}$，则有

$$c'\sqrt{A}(c'\sqrt{A})^T=A \qquad (10\text{-}8\text{-}6)$$

$c'\sqrt{A}$中各元素按下列公式顺序计算

$$c'\sqrt{A_{ij}}=\begin{cases}0 & i>j\\ \left[A_{jj}-\sum\limits_{k=j+1}^{n}(c'\sqrt{A^2_{jk}})\right]^{\frac{1}{2}} & i=j\\ \dfrac{1}{c'\sqrt{A_{jj}}}\left[A_{ij}-\sum\limits_{k=j+1}^{n}c'\sqrt{A_{jk}}c'\sqrt{A_{ik}}\right] & i=j-1,\ j-2,\ \cdots\end{cases}$$

(10-8-7)

$c'\sqrt{A}$的元素是逐列产生的，其顺序如图10-8-2所示。

平方根算法

我们选择比较简单的情况来推导协方差阵平方根算法，设定测量值是标量，状态噪声协方差阵$Q_k=0$，则有一步预报量为

$$\hat{X}_{k/k-1}=\Phi_{k/k-1}\hat{X}_k \tag{10-8-8}$$

$$P_{k/k-1}=\Phi_{k/k-1}P_k\Phi_{k/k-1}^T \tag{10-8-9}$$

把P_k、$P_{k/k-1}$进行平方根分解成下三角阵。

令
$$P_k=S_kS_k^T \tag{10-8-10}$$

$$P_{k/k-1}=S_{k/k-1}S_{k/k-1}^T \tag{10-8-11}$$

把(10-8-9)式写成

$$S_{k/k-1}S_{k/k-1}^T=(\Phi_{k/k-1}S_{k-1})(S_{k-1}^T\Phi_{k/k-1}^T) \tag{10-8-12}$$

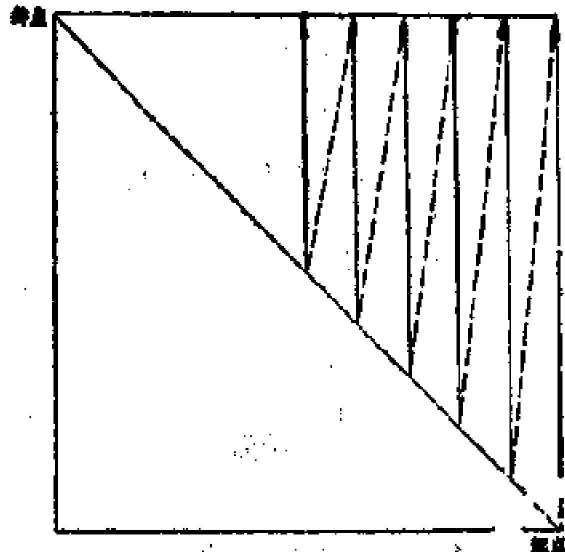


图10-8-2　对A扫描并产生$C=\sqrt{A}$

取下列两式作为预报量

$$\hat{X}_{k/k-1}=\Phi_{k/k-1}\hat{X}_{k-1} \tag{10-8-13}$$

$$S_{k/k-1}=\Phi_{k/k-1}S_{k-1} \tag{10-8-14}$$

在标测量情况下，可得到

$$P_k=P_{k/k-1}-\frac{P_{k/k-1}H_k^TH_kP_{k/k-1}}{H_kP_{k/k-1}H_k^T+R_k} \tag{10-8-15}$$

记
$$a_k=S_{k/k-1}^TH_k^T \tag{10-8-16}$$

$$\frac{1}{b_k}=a_k^Ta_k+R_k \tag{10-8-17}$$

代入(10-8-15)式中

$$S_kS_k^T=S_{k/k-1}[1-b_ka_ka_k^T]S_{k/k-1}^T \tag{10-8-18}$$

其中a_k为$n\times1$维矢量，b_k为标量

再令
$$r=\frac{1}{1+\sqrt{bR}} \tag{10-8-19}$$

则
$$1+b_ka_ka_k^T=(1-b_kra_ka_k^T)(1-b_kra_ka_k^T)^T \tag{10-8-20}$$

代入(10-8-18)式中，得

$$S_k=S_{k/k-1}[1-b_kr_ka_ka_k^T] \tag{10-8-21}$$

相应的增益为

$$K_k=\frac{P_{k/k-1}H_k^T}{H_kP_{k/k-1}H_k^T+R_k}=b_kS_{k/k-1}a_k \tag{10-8-22}$$

把以上结果加以整理，得到协方差平方根算法如下

$$
\left.\begin{aligned}
&\hat{X}_{k+1/k} = \Phi_{k+1/k}\hat{X}_k \\
&S_{k+1/k} = \Phi_{k+1/k} S_k \\
&K_k = b_k S_{k/k-1} a_k \\
&\hat{X}_k = \hat{X}_{k/k-1} + K_k(z_k - H_k \hat{X}_{k/k-1}) \\
&S_k = S_{k/k-1} - r_k K_k a_k^T
\end{aligned}\right\} \qquad (10\text{-}8\text{-}23)
$$

其中 $a_k = S^T_{k/k-1} H_k$

$$b_k = \frac{1}{a_k^T a_k + R_k}$$

$$r_k = \frac{1}{1 + \sqrt{b_k R_k}}$$

例10-8-3　设测量矩阵$H_k = [\frac{1}{3},\ 1]$，$R_k = 4$，

已知　$P_{k/k-1} = \begin{bmatrix} 9 & 9 \\ 9 & 13 \end{bmatrix}$，用平方根求$P_k$。

解：用常规卡尔曼滤波算法，得

$$K_k = P_{k/k-1} H_k^T [H P_{k/k-1} H_k^T + R]^{-1}$$

$$= \frac{1}{20+4}\begin{bmatrix} 12 \\ 16 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{2}{3} \end{bmatrix}$$

$$P_k = P_{k/k-1} - K_k H_k P_{k/k-1}$$

$$= \begin{bmatrix} 9 & 9 \\ 9 & 13 \end{bmatrix} - \begin{bmatrix} \frac{1}{2} \\ \frac{2}{3} \end{bmatrix}[12,\ 16] = \begin{bmatrix} 3 & 1 \\ 1 & \frac{7}{3} \end{bmatrix}$$

再用平方根算法计算，对$P_{k/k-1}$进行平方根分解得　$S_{k/k-1} = \begin{pmatrix} 3 & 0 \\ 3 & 2 \end{pmatrix}$

$$a_k = \begin{pmatrix} 3 & 3 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} \frac{1}{3} \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$$

$$b_k = \frac{1}{4^2 + 2^2 + 4} = \frac{1}{24}$$

$$r_k = \frac{1}{1 + \sqrt{\frac{1}{24} \times 4}} = \frac{1}{1 + \sqrt{\frac{1}{6}}}$$

$$K_k=\frac{1}{24}\begin{bmatrix}3 & 0\\ 3 & 2\end{bmatrix}\begin{bmatrix}4\\ 2\end{bmatrix}=\begin{pmatrix}\frac{1}{2}\\ \frac{2}{3}\end{pmatrix}$$

$$S_k=\begin{bmatrix}3 & 0\\ 3 & 2\end{bmatrix}-\frac{1}{1+\sqrt{\frac{1}{6}}}\begin{bmatrix}\frac{1}{2}\\ \frac{2}{3}\end{bmatrix}[4,\ 2]$$

验证 $S_kS_k^T=\begin{pmatrix}3 & 1\\ 1 & T/3\end{pmatrix}=P_k$。

计算结果表明，平方根算法与常规卡尔曼滤波器算法一致。

对于矢量测量情况以及 $Q_k\neq 0$ 情况，近年来也推导出了协方差平方根算法， 但推导更复杂，计算量又大，应用不多，这里从略。

为了克服数值发散，除了上述协方差平方根算法外，还有一种协方差 U-D 因子分解算法。在此算法中，不需要把协方差阵分解成平方根因子，而是把它分成下列形式

$$P=UDU^{-1}$$

其中D为对角线阵，U为上三角阵，且U的对角线元素均为1。当P为 $n\times n$ 半正定矩阵时，这种分解一定成立，由于其推导亦较繁，故这里从略。

第四篇 谱 估 计

第十一章 线性功率谱估计

在平稳随机过程的信号处理中，中心问题之一是估计一个时间序列的功率谱，而这个时间序列是在一个有限持续时间内的观测值。无疑，功率谱估计具有很重要的实际意义。例如，在雷达和声纳检测中，为寻找掩埋于环境干扰中的信号，需要对接收的样本进行谱估计，以提取有用信息；对于机械系统运动所产生的噪声进行谱分析，有助于给设计者获得有效的设计参数等等。

功率谱估计的方法可分为两大类：线性谱估计方法和非线性谱估计方法。线性谱估计方法或常规方法只对有效时间序列进行线性运算。常用的是由Blackman和Tukey提出的相关函数法（BT法）和周期图法（或称FFT法）。不论哪一种方法均有一个共同的缺点：有假的谱分量。这是因为它们均涉及到一个“窗”函数的问题。这种窗是人为的，与时间序列本身的特性无关。后面的分析将会说明这一点。非线性谱分析方法与线性谱分析方法不同，它避免了人为“窗”的问题。对于短数据样本，非线性谱估计方法显示出了它的优点，分辨力优于常规方法。

这一章叙述线性功率谱估的基本原理、关于非线性功率谱估计方法将在下一章进行讨论。

§11-1 相关函数法

我们知道，一个平稳时间序列 $x(n)$ 的功率谱定义为它的相关函数的付里叶变换

$$S_x(\omega) = T_s \sum_{m=-\infty}^{+\infty} \Phi_{xx}(m) e^{-j\omega mT_s} \tag{11-1-1}$$

因而有

$$\Phi_{xx}(m) = \frac{1}{2\pi}\int_{-\pi/T_s}^{\pi/T_s} S_x(\omega) e^{j\omega mT_s} \tag{11-1-2}$$

这里$\Phi_{xx}(m)$为时间序列$x(n)$在时延值为mT_s时的自相关函数，定义如下：

$$\Phi_{xx}(m) = \Phi_{xx}(mT_s) = E[x_n x_{n+m}] \tag{11-1-3}$$

当过程具有各态历经特性时，可用时间平均代替统计平均，此时有

$$r_{xx}(m) = \Phi_{xx}(m) = \lim_{N\to\infty}\frac{1}{2N+1}\sum_{n=-N}^{N} x(n)x(n+m) \tag{11-1-4}$$

因此，一个离散平稳时间序列 $x(n)$ 的功率谱原则上可由(11-1-4) 及 (11-1-1)式求得。由于实际上时间序列总是有限的，因而实际上所得的谱只能是理论值的一种估计。

对于有限长N的实因果序列，其相关函数的估值为

$$\hat{r}_{xx}(m) = \frac{1}{N}\sum_{n=0}^{N-1-|m|} x(n)x(n+|m|) \qquad |m| \leqslant N-1 \tag{11-1-5}$$

因而应序列的功率谱密度为

$$\hat{S}_x(\omega) = T_s \sum_{m=-(N-1)}^{N-1} \hat{r}_{xx}(m) e^{-j\omega m T_s} \tag{11-1-6}$$

不失一般性，今后令 $T_s = 1$。

现在来分析(11-1-5)式表示的相关函数估值的性质。

$\hat{r}_{xx}(m)$ 的均值为

$$E[\hat{r}_{xx}(m)] = \frac{1}{N}\sum_{n=0}^{N-|m|-1} E[x(n)x(n+|m|)]$$

$$= \frac{N-|m|}{N} r_{xx}(m) \tag{11-1-7}$$

估值的偏量为

$$B = r_{xx}(m) - E[\hat{r}_{xx}(m)] = \frac{|m|}{N} r_{xx}(m) \tag{11-1-8}$$

当$N\to\infty$时$B\to 0$，因此可知(11-1-5) 式给出的自相关函数估计是渐近无偏的。

$\hat{r}_{xx}(m)$的方差为

$$\begin{aligned}\mathrm{var}[\hat{r}_{xx}(m)] &= E[\{(\hat{r}_{xx}(m) - E[\hat{r}_{xx}(m)]\}^2] \\ &= E[\hat{r}_{xx}^2(m)] - (E[\hat{r}_{xx}(m)])^2\end{aligned} \tag{11-1-9}$$

上式第一项为

$$E[\hat{r}_{xx}^2(m)] = E\left[\frac{1}{N^2}\sum_{n=0}^{N-|m|-1}\sum_{k=0}^{N-|m|-1} E[x(n)x(n+|m|)\right.$$

$$\left.\times x(k)x(k+|m|)]\right]$$

为简化分析，设 $x(n)$ 是均值为零的高斯随机序列。由概率论可知有下列关系式

$$\begin{aligned}E[x_1x_2x_3x_4] = E[x_1x_2]E[x_3x_4] + E[x_1x_3]E[x_2x_4] \\ + E[x_1x_4]E[x_2x_3]\end{aligned} \tag{11-1-10}$$

因此可以得到

$$\begin{aligned}&E[x(n)x(n+|m|)\cdot x(k)x(k+|m|)] \\ &= E[x(n)x(n+|m|)]E[x(k)x(k+|m|)] \\ &\quad + E[x(n)x(k)]E[x(n+|m|)x(k+|m|)] \\ &\quad + E[x(n)x(k+|m|)]E[x(k)x(n+|m|)] \\ &= r_{xx}^2(m) + r_{xx}^2(n-k) + r_{xx}(n-k+m)r_{xx}(n-k-m)\end{aligned} \tag{11-1-11}$$

将(11-1-11)式代入(11-1-9)式，得到

$$\begin{aligned}\mathrm{var}[\hat{r}_{xx}(m)] = \frac{1}{N^2}\sum_{n=0}^{N-|m|-1}\sum_{k=0}^{N-|m|-1} [r_{xx}^2(n-k) \\ + r_{xx}(n-k+m)r_{xx}(n-k-m)]\end{aligned} \tag{11-1-12}$$

进一步化简后不难得到

$$\mathrm{var}[\hat{r}_{xx}(m)]=\frac{1}{N}\sum_{l=-(N-1-|m|)}^{N-1-|m|}\left[1-\frac{|m|+|l|}{N}\right]$$

$$\times[r_{xx}^2(l)+r_{xx}(l+m)r_{xx}(l-m)] \qquad (11\text{-}1\text{-}13)$$

容易得知，当$N\to\infty$时，$\mathrm{var}[\hat{r}_{xx}(m)]\to 0$，因此(11-1-5)式给出的自相关估计是渐近无偏的一致估计。

由上分析可知，采用相关法估计功率谱的步骤是首先由$x(n)$按(11-1-5)式估计相关序列，然后将结果求一次离散付里叶变换，得到(11-1-6)式的$\hat{S}_x(\omega)$。但直接计算相关函数运算量较大，因而常采用FFT算法进行计算。

利用循环相关计算，可以得到长度为N的序列的N点线性相关函数。这只要将序列长度补零扩大到$2N-1$点，先计算$2N-1$点的FFT，再计算循环相关函数，即

$$\hat{r}_{xx}(m)=\frac{1}{N}\sum_{n=0}^{2N-2}[\tilde{x}(n)\,\tilde{x}(n+m)]R_N(m)$$

$$=\frac{1}{N}\mathrm{IDFT}[X(k)X^*(k)] \qquad (11\text{-}1\text{-}14)$$

其中

$$R_N(m)=\begin{cases}1 & 0\leqslant m\leqslant N-1\\ 0 & 其它\end{cases} \qquad (11\text{-}1\text{-}15)$$

$\tilde{x}(n)$为N点序列补零扩展到$2N-1$点之后的序列，$X(k)$为$\tilde{x}(n)$的DFT，$*$表示复共轭。

对于长序列的相关函数和功率谱密度函数，可用分段技术，其中有重迭相加法及重迭保留法，读者可参阅有关文献[21, 24]。

§11-2 周 期 图 法

在估计随机信号的谱密度的线性方法中，除了以自相关函数的离散付里叶变换来获得之外，还可以直接从随机信号序列本身的付里叶变换来获得。这就是所谓周期图法。周期图定义为序列$x(n)$的$2N-1$点自相关估值$\hat{r}_{xx}(m)$的付里叶变换

$$I_N(\omega)=\hat{S}_x(\omega)=\sum_{m=-(N-1)}^{N-1}\hat{r}_{xx}(m)e^{-j\omega m}$$

可以证明，周期图与有限长度序列$x(n)$的付里叶变换有如下关系

$$I_N(\omega)=\frac{1}{N}|X(e^{j\omega})|^2 \qquad (11\text{-}2\text{-}1)$$

先把$x(n)$补零至$2N-1$点长度，再求出

$$X(e^{j\omega})=\sum_{n=0}^{(2N-2)}x(n)e^{-j\omega n}$$

是序列$x(n)$的付里叶变换。因为$X(e^{j\omega})$具有周期性，因而周期图也具有周期性。利用周期图估计序列$x(n)$的功率谱时，只要对序列作DFT运算，然后取其绝对值平方，再按序列长度平均。因此，用FFT方法很容易估计一个平稳随机序列的功率谱。

我们感兴趣的是这种功率谱估计的性能，与在相关法中讨论类似，研究一下这种估计的均值与方差。

$I_N(\omega)$的均值为

$$E[I_N(\omega)]=\sum_{m=-(N-1)}^{N-1}E[\widehat{r_{xx}}(m)]e^{-j\omega m} \tag{11-2-2}$$

将上节的(11-1-7)式表示的$E[\widehat{r_{xx}}(m)]$代入(11-2-2)式得到

$$E[I_N(\omega)]=\sum_{m=-(N-1)}^{N-1}\left(\frac{N-|m|}{N}\right)r_{xx}(m)e^{-j\omega m} \tag{11-2-3}$$

由于求和限有限，以及由于$(N-|m|)/N$因子的作用，$E[I_N(\omega)]$不等于相关函数$r_{xx}(m)$的付里叶变换，因而周期图是功率谱$S_x(\omega)$的有偏估计。

(11-2-3)式可解释为自相关序列$r_{xx}(m)$加窗后的付里叶变换。这里窗函数为

$$w_B(m)=\begin{cases}\dfrac{N-|m|}{N} & |m|<N\\ 0 & \text{其它}\end{cases} \tag{11-2-4}$$

是一个三角形窗，或称巴特利特窗。因而$E[I_N(\omega)]$可写成

$$E[I_N(\omega)]=\sum_{m=-(N-1)}^{N-1}w_B(m)r_{xx}(m)e^{-j\omega m} \tag{11-2-5}$$

现在来研究一下周期图的方差。

设$x(n)$是均值为零的白高斯序列，周期图(11-2-1)可表示为

$$I_N(\omega)=\frac{1}{N}\sum_{l=0}^{N-1}\sum_{r=0}^{N-1}x(l)x(r)e^{-j\omega l}\cdot e^{-j\omega r} \tag{11-2-6}$$

为计算$I_N(\omega)$的方差，首先计算在两个频率ω_1和ω_2处的$I_N(\omega)$的协方差

$$\begin{aligned}\operatorname{cov}[I_N(\omega_1),\ I_N(\omega_2)]\\ &=E[(I_N(\omega_1)-E(I_N(\omega_1)))(I_N(\omega_2)-E(I_N(\omega_2)))]\\ &=E[I_N(\omega_1)I_N(\omega_2)]-E[I_N(\omega_1)]E[I_N(\omega_2)]\end{aligned} \tag{11-2-7}$$

上式右边第一项为

$$\begin{aligned}&E[I_N(\omega_1)I_N(\omega_2)]\\ &=\frac{1}{N^2}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}E[x(k)x(l)x(m)x(n)]e^{j[(k-l)\omega_1+(m-n)\omega_2]}\end{aligned} \tag{11-2-8}$$

对于白高斯过程，注意利用(11-1-10)式，有

$$E[x(n)x(n+m)]=\sigma_x^2\delta(m)$$

$$E[x(k)x(l)x(m)x(n)]=\begin{cases}\sigma_x^4, & k=l,\ m=n;\ 或k=m,\\ & l=n;\ 或\ k=n, l=m\\ 0, & 其它\end{cases} \tag{11-2-9}$$

将(11-2-9)式代入(11-2-8)式得

$$E[I_N(\omega_1)I_N(\omega_2)]=\frac{\sigma_x^4}{N^2}\left\{N^2+\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}e^{j(m-n)(\omega_1+\omega_2)}\right.$$

$$+\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}e^{j(n-m)(\omega_1-\omega_2)}\Big\} \tag{11-2-10}$$

将上式中二重求和化为两个和式的积，经过简化不难得到

$$E[I_N(\omega_1)I_N(\omega_2)]=\sigma_x^4\left\{1+\left(\frac{\sin[(\omega_1+\omega_2)N/2]}{N\sin[(\omega_1+\omega_2)/2]}\right)^2\right.$$

$$\left.+\left(\frac{\sin[(\omega_1-\omega_2)N/2]}{N\sin[(\omega_1-\omega_2)/2]}\right)^2\right\} \tag{11-2-11}$$

将(11-2-11)式代入 (11-2-7)式，并注意到 $E[I_N(\omega_1)]=E[I_N(\omega_2)]=\sigma_x^2$，可得到周期图的协方差为

$$\mathrm{cov}[I_N(\omega_1),I_N(\omega_2)]=\sigma_x^4\left\{\left(\frac{\sin[(\omega_1+\omega_2)N/2]}{N\sin[(\omega_1+\omega_2)/2]}\right)^2\right.$$

$$\left.+\left(\frac{\sin[(\omega_1-\omega_2)N/2]}{N\sin[(\omega_1-\omega_2)/2]}\right)^2\right\} \tag{11-2-12}$$

因此，在$\omega=\omega_1=\omega_2$这一特定频率谱估值的方差为

$$\mathrm{var}[I_N(\omega)]=\mathrm{cov}[I_N(\omega),I_N(\omega)]$$

$$=\sigma_x^4\left[1+\left(\frac{\sin N\omega}{N\sin\omega}\right)^2\right] \tag{11-2-13}$$

当$N\to\infty$时，$I_N(\omega)$的方差不趋于零。因此周期图不是功率谱的一致估计，不论N 如何取，其方差均为σ_x^4的量级。随着N的增大，方差趋于常数。

§11-3　谱估值的平滑

上节的讨论表明，随N的增大，周期图的特性并不令人满意。为了减少周期图功率谱估计的方差，可以采用几种方法。本节将介绍如何使用周期图才能得到谱的一致估计。

一、平均周期图法——巴特利特(Bartlett)法

将数据列$x(n)$ 分成K段，每一段有M个样值，因而$N=KM$。即构成如下的序列段

$$x_i(n)=x(n+(i-1)M),\quad 0\leqslant n\leqslant M-1,\quad 1\leqslant i\leqslant K$$

计算K个周期图

$$I_{M_i}=\frac{1}{M}\left|\sum_{n=0}^{M-1}x_i(n)e^{-j\omega n}\right|^2\qquad 1\leqslant i\leqslant K \tag{11-3-1}$$

若$m>M$时，$r_{xx}(m)$很小，此时可设诸周期图$I_{M_i}(\omega)$彼此独立。定义功率谱估值为

$$\hat{S}_x(\omega)=\frac{1}{K}\sum_{i=1}^{K}I_{M_i}(\omega) \tag{11-3-2}$$

这一方法首先由巴特利特提出，故称为巴特利方法。该估值的期望值为

$$E[\hat{S}_x(\omega)]=\frac{1}{K}\sum_{i=1}^{K}E[I_{M_i}(\omega)]=E[I_M(\omega)] \tag{11-3-3}$$

注意到上节(11-2-5)式，可将周期图的均值写为窗谱$W_B(e^{j\omega})$与真谱的卷积

$$E[I_N(\omega)]=\frac{1}{2\pi}\int_{-\pi}^{\pi}S_x(\theta)W_B(e^{j(\omega-\theta)})d\theta \tag{11-3-4}$$

其中

$$W_B(e^{j\omega})=\frac{1}{N}\left(\frac{\sin\frac{N}{2}\omega}{\sin\frac{\omega}{2}}\right)^2 \tag{11-3-5}$$

将(11-3-5)式代入(11-3-4)式，并将N换为M得到

$$E[\hat{S}_x(\omega)]=E[I_{M_i}(\omega)]$$

$$=\frac{1}{2\pi M}\int_{-\pi}^{\pi}S_x(\theta)\left[\frac{\sin\frac{M}{2}(\omega-\theta)}{\sin\frac{1}{2}(\omega-\theta)}\right]^2d\theta \tag{11-3-6}$$

因此，用平均周期图法得到的功率谱估计的期望值等于真实功率谱$S_x(\omega)$与三角形窗谱函数的卷积，且相当于某个长度为M的样本序列的周期图的期望值，因而平均周期图也是一个有偏估计。

现在来分析平均周期图的方差。

因为已假定(11-3-2)式中求平均的K个周期图是统计独立的，因此功率谱估计的方差为

$$\mathrm{var}[\hat{S}_x(\omega)]=\frac{1}{K}\mathrm{var}[I_M(\omega)] \tag{11-3-7}$$

将(11-2-13)式的结果代入，并将N换为M，得到

$$\mathrm{var}[\hat{S}_x(\omega)]=\frac{1}{K}\sigma_x^4\left[1+\left(\frac{\sin\omega M}{M\sin\omega}\right)^2\right]$$

$$\approx\frac{1}{K}S_x^2(\omega)\left[1+\left(\frac{\sin\omega M}{M\sin\omega}\right)^2\right] \tag{11-3-8}$$

可见方差$\mathrm{var}\,\hat{S}_x(\omega)$是与被平均的周期图的数目成反比的，并且随着$K$的增大，方差趋近于零。因此平均周期图法的功率谱估计是一致估计。

若数据总长度N不变，仅增加分段数目K，则由于M的减少，相应的功率谱主瓣加宽，分辨力下降。必须在分辨力与方差之间折衷处理。增加数据列的长度，可以兼顾分辨力与方差的要求。

二、平滑周期图法(窗口函数法)

由上面的讨论可以看到，平均周期图法由于采用了较短的记录长度，而降低了分辨力。减小谱估计的方差是以增加偏量和降低分辨力为代价的。用一个合适的谱窗和周期图进行卷积，从而对周期图进行平滑是本节介绍的另一种降低估计方差的方法。

若$\overline{S}_x(\omega)$是已平滑了的周期图，$W(e^{j\omega})$为谱窗函数，则有

$$\overline{S}_x(\omega)=\frac{1}{2\pi}\int_{-\pi}^{\pi}I_N(\theta)W(e^{j(\omega-\theta)})d\theta \tag{11-3-9}$$

因为周期图是$\hat{r}_{xx}(m)$的付里叶变换，而谱窗函数$W(e^{j\omega})$的逆付里叶变换

$$w(m)=\frac{1}{2\pi}\int_{-\pi}^{\pi}W(e^{j\omega})e^{j\omega m}d\omega \qquad (11\text{-}3\text{-}10)$$

是长度为$2N-1$的有限时宽窗函数序列。根据频率域卷积定理，(11-3-9)式又可写成

$$\overline{S}_x(\omega)=\sum_{m=-(M-1)}^{M-1}\hat{r}_{xx}(m)w(m)e^{-j\omega m} \qquad (11\text{-}3\text{-}11)$$

由(11-3-9)及(11-3-11)式可以得知，已平滑了的功率谱$\overline{S}_x(\omega)$可认为是由周期图$I_N(\omega)$或具有自相关估值$\hat{r}_{xx}(m)$的信号通过具有冲激响应序列$w(m)$或频率响应$W(e^{j\omega})$的滤波器而得到的。由于窗口函数具有低通特性，因而对周期图有平滑作用。

现在具体分析一下平滑周期图法谱估计的均值与方差。

均值为

$$E[\overline{S}_x(\omega)]=\frac{1}{2\pi}\int_{-\pi}^{\pi}E[I_N(\theta)]W(e^{j(\omega-\theta)})d\theta \qquad (11\text{-}3\text{-}12)$$

注意到(11-3-4)式，容易看出$E[\overline{S}_x(\omega)]$是$W_B(e^{j\omega})$和$W(e^{j\omega})$两谱窗函数与真谱$S_x(\omega)$的频域卷积。在时域上的表示式为

$$E[\overline{S}_x(\omega)]=\sum_{m=-(M-1)}^{M-1}r_{xx}(m)w_B(m)w(m)e^{-j\omega m} \qquad (11\text{-}3\text{-}13)$$

式中

$$w_B(m)=1-\frac{|m|}{N} \qquad |m|\leqslant N$$

若M比N小，则$W(e^{j\omega})$将比$W_B(e^{j\omega})$宽，上式中窄窗函数序列$w(m)$起主要作用，卷积公式可近似写成

$$E[\overline{S}_x(\omega)]\approx\frac{1}{2\pi}\int_{-\pi}^{\pi}S_x(\theta)W(e^{j(\omega-\theta)})d\theta \qquad (11\text{-}3\text{-}14)$$

由(11-3-12)和(11-3-14)式可见，功率谱窗函数$W(e^{j\omega})$宽度增大，功率谱将得到进一步平滑，但功率谱主瓣加宽，使谱估计量频率分辨力降低。当记录长度N变大时，可以加大窗函数$w(m)$的宽度，使$W(e^{j\omega})$比$S_x(\omega)$的宽度窄，从而有

$$\lim_{M\to\infty}E[\overline{S}_x(\omega)]=S_x(\omega)\ \frac{1}{2\pi}\int_{-\pi}^{\pi}W(e^{j\omega})d\omega \qquad (11\text{-}3\text{-}15)$$

如果满足条件

$$w(0)=\frac{1}{2\pi}\int_{-\pi}^{\pi}W(e^{j\omega})d\omega=1 \qquad (11\text{-}3\text{-}16)$$

则有

$$\lim_{M\to\infty}E[\overline{S}_x(\omega)]=S_x(\omega) \qquad (11\text{-}3\text{-}17)$$

它表明平滑周期图功率谱估计是渐近无偏的。

现在来分析估计的方差。

与在§11-2中一样，设$x(n)$是零均值白高斯序列，仿照该节的推导过程，先计算

$$\mathrm{cov}[\overline{S}_x(\omega_1),\ \overline{S}_x(\omega_1)]=E[\{\overline{S}_x(\omega_1)-E[\overline{S}_x(\omega_1)]\}$$

$$\cdot\{\overline{S}_x(\omega_2) - E[\overline{S}_x(\omega_2)]\}] \qquad (11\text{-}3\text{-}18)$$

由(11-3-9)和(11-3-12)式得到

$$\overline{S}_x(\omega) - E[\overline{S}_x(\omega)]$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}[I_N(\theta) - E[I_N(\theta)]]W(e^{j(\omega-\theta)})d\theta$$

因此(11-3-18)可写为

$$\operatorname{cov}[\overline{S}_x(\omega_1), \overline{S}_x(\omega_2)] \approx \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}W(e^{j(\omega_1-\theta)})$$

$$\times W(e^{j(\omega_2-\varphi)})\operatorname{cov}[I_N(\theta), I_N(\varphi)]d\theta d\varphi \qquad (11\text{-}3\text{-}19)$$

注意到§11-2中的(11-2-12)式，但采用非白高斯序列周期图方差的更一般表示式[24]

$$\operatorname{cov}[I_N(\theta), I_N(\varphi)]$$

$$\approx S_x(\theta)S_x(\varphi)\left[\left(\frac{\sin[(\theta+\varphi)N/2]}{N\sin[(\theta+\varphi)/2]}\right)^2\right.$$

$$\left. + \left(\frac{\sin[(\theta-\varphi)N/2]}{N\sin[(\theta-\varphi)/2]}\right)^2\right] \qquad (11\text{-}3\text{-}20)$$

假定上式中两个正弦函数的平方项比起 $S_x(\theta)$ 的变化快很多，即在N很大时它们分别高度集中在$\theta = -\varphi$和$\theta = \varphi$附近。首先对θ进行积分，并注意利用关系式

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\sin(\theta N/2)}{N\sin(\theta/2)}\right)^2 d\theta = \frac{1}{N}$$

可近似得到

$$\operatorname{cov}[\overline{S}_x(\omega_1), \overline{S}_x(\omega_2)]$$

$$\approx \frac{1}{2\pi N}\int_{-\pi}^{\pi}S_x^2(\varphi)W(e^{j(\omega_2-\varphi)})[W(e^{j(\omega_1+\varphi)}) + e^{j(\omega_1-\varphi)}]d\varphi \qquad (11\text{-}3\text{-}21)$$

进一步假设谱窗函数$W(e^{j\omega})$足够窄，以致相距较远的两个谱窗函数之乘积$W(e^{j(\omega_1+\varphi)})\times W(e^{j(\omega_1-\varphi)})$很小，可以忽略，因此(11-3-21)式变成

$$\operatorname{cov}[\overline{S}_x(\omega_1), \overline{S}_x(\omega_2)]$$

$$\approx \frac{1}{2\pi N}\int_{-\pi}^{\pi}\overline{S}_x(\omega)W(e^{j(\omega_2-\varphi)})W(e^{j(\omega_1-\varphi)})d\varphi \qquad (11\text{-}3\text{-}22)$$

从而由(11-3-22)式可得 $\omega_1 = \omega_2 = \omega$ 时的协方差，即谱估计$\overline{S}_x(\omega)$的方差为

$$\operatorname{var}[\overline{S}_x(\omega)] \approx \frac{1}{2\pi N}\int_{-\pi}^{\pi}S_x^2(\varphi)W^2(e^{j(\omega-\varphi)})d\varphi \qquad (11\text{-}3\text{-}23)$$

假定窗谱函数$W(e^{j\omega})$相对于序列的真谱$S_x(\omega)$而言窄得多，亦即使窗函数$w(m)$长度足够大，保证满意的频谱分辨力。此时，(11-3-23)式可进一步写成

$$\operatorname{var}[\overline{S}_x(\omega)] \approx \frac{1}{N}S_x^2(\omega)\cdot\frac{1}{2\pi}\int_{-\pi}^{\pi}W^2(e^{j\omega})d\varphi \qquad (11\text{-}3\text{-}24)$$

注意到窗函数$w(m) = w(-m)$，以及有

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} W^2(e^{j\varphi})d\varphi = \sum_{m=-(M-1)}^{M-1} w^2(m)$$

则(11-3-24)式可以改写成

$$\text{var}[\overline{S}_x(\omega)] \approx \left(\frac{1}{N}\sum_{m=-(M-1)}^{M-1} w^2(m)\right) S_x^2(\omega) \tag{11-3-25}$$

由(11-3-25)式可知，使平滑周期图谱估计的方差小的关键是使$\frac{1}{N}\sum_{m=-(M-1)}^{M-1} w^2(m)$ $\left(或\frac{1}{2\pi N}\int_{-\pi}^{\pi} W^2(e^{j\omega})d\omega\right)$尽量小。(11-3-25)式成立的条件是：谱窗函数$W(e^{j\omega})$的宽度比真谱$S_x(\omega)$窄(即窗函数$\omega(m)$的长度$2M-1$要足够长)，但比$[\sin(\omega N/2)/\sin(\omega/2)]^2$项宽度要宽。

比较周期图的方差和平滑周期图的方差可知，由于加窗处理，使方差减少。加窗处理后方差改善因子为$\frac{1}{N}\sum_{m=-(M-1)}^{M-1} w^2(m)$。

三、韦尔奇(Welch)法——平滑周期的平均法[25]

这一方法是对巴特利特法的一种修正，特别适用于FFT直接计算功率谱估值。它与巴特利特法的不同之处是在计算周期图之前将窗函数$w(m)$直接加到各数据段上去。

将长为N的信号序列分成K段，每段长为M，每段序列仍表示为

$$x_i(n) = x(n+(i-1)M),\quad 0\leqslant n\leqslant M-1,\quad 1\leqslant i\leqslant K$$

先用窗函数$w(m)$对其加权，得到K个修正的周期图为

$$I_{Mi}^w(\omega) = \frac{1}{MU}\left|\sum_{n=0}^{M-1} x_i(n)w(n)e^{-j\omega n}\right|^2,\quad i=1,2,\cdots,k \tag{11-3-26}$$

其中U表示窗口序列的平均功率。

$$U = \frac{1}{M}\sum_{n=0}^{M-1} w^2(m) \tag{11-3-27}$$

按本节(11-3-2)式的形式，可得功率谱估计为

$$\hat{S}_x^w(\omega) = \frac{1}{K}\sum_{i=1}^{K} I_{Mi}^w(\omega) \tag{11-3-28}$$

由(11-3-26)式可知，不论采用何种窗函数，功率谱估计$\hat{S}_x^w(\omega)$均为正实函数。

现在研究一下谱估计的性质。首先看一下期望值。

经加权后的序列为$x_i(n)w(n)$，它的付里叶变换为

$$X_i(e^{j\omega}) = \sum_{n=0}^{M-1} x_i(n)w(n)e^{-j\omega n} \tag{11-3-29}$$

此时K个修正的周期图成为

$$I_{Mi}^w = \frac{1}{MU}|X_i(e^{j\omega})|^2 \tag{11-3-30}$$

由(11-3-28)式，可写出功率谱估计的期望值为

$$E[\hat{S}_x^s(\omega)] = \frac{1}{K}\sum_{i=1}^{K} E[I_{Mi}^s(\omega)]$$

$$= \frac{1}{K}\sum_{i=1}^{K} E\left[\frac{1}{MU}|X_i(e^{j\omega})|^2\right]$$

$$= E\left[\frac{1}{MU}|X_i(e^{j\omega})|^2\right] \tag{11-3-31}$$

记 $Y_i(\omega)=\frac{1}{\sqrt{MU}}X_i(e^{j\omega})$，利用频率卷积定理可以得到

$$E[Y_i(\omega)Y_i^*(\omega)]=\frac{1}{2\pi}\int_{-\pi}^{\pi}S_x(\theta)W(e^{j(\omega-\theta)})d\theta \tag{11-3-32}$$

其中

$$W(e^{j\omega})=\frac{1}{MU}\left|\sum_{n=0}^{M-1}w(n)e^{-j\omega n}\right|^2 \tag{11-3-33}$$

于是$\hat{S}_x^s(\omega)$的期望值可写为

$$E[\hat{S}_x^s(\omega)]=E[Y_i(\omega)Y_i^*(\omega)]$$

$$=\frac{1}{2\pi}\int_{-\pi}^{\pi}S_x(\theta)W(e^{j(\omega-\theta)})d\theta \tag{11-3-34}$$

为使估计量$\hat{S}_x^s(\omega)$渐近无偏，归一化因子U是必不可少的。

与平均周期图中的分析类似，易于得到平均平滑周期图的方差为

$$\mathrm{var}[\hat{S}_x^s(\omega)]\approx\frac{1}{K}S_x^2(\omega) \tag{11-3-35}$$

可见，该估计是一致估计。为使K足够大，要求信号序列尽可能长。

第十二章　非线性功率谱估计

上一章讨论了线性功率谱估计方法，本章将讨论近年来发展起来的非线性功率谱估计方法。

正如上一章所述，线性谱估计方法的缺点是可能导致假谱，原因是引入了人为的窗函数。如果数据列很短，问题就更加严重，以致找不到一个合适的窗函数来辨别出所感兴趣的频率分量。非线性谱估计方法克服了窗函数的缺点。之所以称为非线性方法是因为它们是根据数据样本进行运算的。

自1938年瓦尔德（Wold）提出序列分解的基本定理以来，序列的自回归（AR）模型虽然有了相当多的应用，但在谱分析方面并未引起人们广泛的注意。直到 1967 年伯格（Burg）第一次提出最大熵谱的概念，随后Van de Bos在1971 年指出了最大熵谱分析方法与AR谱分析方法的等价性之后才真正认识了AR分析的价值。1969年由Capon提出的最大似然方法，首先用于线阵的空间波数谱分析，Lacoss 将这种方法应用于时间序列的谱分析。随后伯格又证明了最大似然谱是经平滑的最大熵谱的型式。近年来，Shore 在最大熵谱分析的基础上提出了最小交叉熵的概念，并提出了最大熵谱只是最小交叉熵方法的一种特例。由于最小交叉熵方法利用了某些先验信息，因而所得到的谱比最大熵谱有更好的质量。此外滑动自回归（ARMA）模型也是近年来人们感兴趣的方法。但由于最小交叉熵谱以及 ARMA 谱方法计算的复杂性。使得它们还未得到如最大熵谱方法的广泛应用。

§12-1　最大熵谱估计

一、熵的概念——有限高斯过程的熵率

在讨论最大熵谱分析方法之前，有必要首先研究一下平稳过程的熵。

根据山农信息理论，若一个系统传输M个不同的消息x_1，x_2，…，x_M，其概率分别为p_1，p_2，…，p_M，消息的平均信息量

$$H=\sum_{j=1}^{M}p_j\log_2\frac{1}{p_j}=-\sum_{j=1}^{M}p_j\log_2 p_j \tag{12-1-1}$$

定义为熵。式中对数亦可取自然对数，只差一常数。熵表明事件出现的不确定性，这个数量愈大，则事件愈不确定，亦即消息的信息量愈大。

一个平稳随机序列$x(n)$的熵可用下式表示

$$H=-c\int_{-\infty}^{\infty}p(x)\ln[p(x)]dx \tag{12-1-2}$$

其中$p(x)$是序列$x(n)$（$n=1,\cdots,N$）的N维联合概率密度函数。对于高斯型序列，联合概率密度函数为

$$p(x)=\frac{1}{[(2\pi)^N\det(C_x)]^{1/2}}\exp\left[-\frac{1}{2}(x-\mu_x)^H C_x^{-1}(x-\mu_x)\right] \tag{12-1-3}$$

其中C_x为协方差矩阵，其元素为

$$c_{ij}=E[(x_i-\mu_x)(x_j-\mu_x)^*]$$

以及"H"表示复共轭转置，$\mu_x=E[x_i]=E[x_j]$，μ_x表示均值向量。若过程均值为零，协方差矩阵就变为自相关矩阵。将(12-1-3)式代入(12-1-2)式得

$$H=\frac{1}{2}\ln[\det(C_x)] \tag{12-1-4a}$$

当过程均值为零时，则有

$$H=\frac{1}{2}\ln[\det(\Phi_x)] \tag{12-1-4b}$$

式中 Φ_x 为自相关矩阵。当过程无限长时，熵将发散，必须用熵率作为信息量的度量，它定义为

$$h=\lim_{M\to\infty}\frac{H}{M+1}=\lim_{M\to\infty}\frac{1}{2}\ln[\det(\Phi_x)]^{1/M+1} \tag{12-1-5}$$

这一行列式的极限与时间序列$x(n)$的谱$S_x(f)$之间的关系为

$$\lim_{M\to\infty}[\det(\Phi_x)]^{\frac{1}{M+1}}=2B\cdot\exp\left\{\frac{1}{2B}\int_{-B}^{B}\ln[S_x(f)]df\right\} \tag{12-1-6}$$

其中$2B$为过程的带宽，即$-B\leqslant f\leqslant B$。

利用(12-1-6)式，可将(12-1-5)式写成

$$h=\frac{1}{2}\ln(2B)+\frac{1}{4B}\int_{-B}^{B}\ln[S_x(f)]df \tag{12-1-7a}$$

或忽略常数项，有

$$h=\frac{1}{4B}\int_{-B}^{B}\ln[S_x(f)]df \tag{12-1-7b}$$

上式描述了过程 $x(n)$的谱与熵之间的关系，它适用于任何平稳时间序列。

二、最大熵谱分析

现在讨论如何利用熵率确定正实数谱的问题。具体说，即按这样一个准则来估计功率谱：该谱所对应的相关函数与已知$2M+1$个相关值一致，而对于未知相关值的区域，熵率最大。

一个平稳随机过程的功率谱密度与自相关函数 Φ_x 有下列关系

$$S_x(f)=\frac{1}{2B}\sum_{k=-\infty}^{+\infty}r_{xx}(k)e^{-j2\pi fkT} \tag{12-1-8}$$

其中$r_{xx}(k)$是相关矩阵Φ_x的元素。将(12-1-8)式代入(12-1-7b)式，并令

$$\frac{\partial h}{\partial r_{xx}(k)}=0,\quad |k|\geqslant M+1$$

不难得到

$$\int_{-B}^{B}\frac{\exp(-2\pi kfT)}{\hat{S}_x(f)}df=0,\quad |k|\geqslant M+1 \tag{12-1-9}$$

该式意味着功率谱的估值的倒数可表示为一个截断的付里叶级数，即

$$\frac{1}{\widehat{S}_x(f)}=\sum_{n=-M}^{M}c_n\exp(-j2\pi nfT) \tag{12-1-10}$$

其中c_n为付里叶展开式的系数，为保证$\widehat{S}_x(f)$为正实数，有$c_{-n}=c_n^*$

下一步是求出系数c_k，它们必须使$\widehat{S}_x(f)$满足与已知自相关函数值一致的条件。亦即

$$r_{xx}(k)=\int_{-B}^{B}\widehat{S}_x(f)\exp(j2\pi kfT)df,\quad |k|\leqslant M \tag{12-1-11}$$

将(12-1-10)式代入(12-1-11)式，并用Z变换(令$Z=\exp(j2\pi fT)$)，得到

$$r_{xx}(k)=\frac{2B}{2\pi Tj}\oint\frac{Z^{k-1}}{\sum\limits_{n=-M}^{M}c_nZ^{-n}}dZ,\quad 0\leqslant k\leqslant M \tag{12-1-12}$$

上式围线积分沿单位圆逆针方向进行。由于系数$\{c_n\}$的埃尔密特性质，分母的和式可表为两个多项式之积，即

$$\sum_{n=-M}^{M}c_nZ^{-n}=G_M(Z)G_M^*(1/Z^*) \tag{12-1-13}$$

其中

$$G_M(Z)=\sum_{n=0}^{M}g_nZ^{-n}$$

$$G_M^*(1/Z^*)=\sum_{n=0}^{M}g_n^*Z^n \tag{12-1-14}$$

第一个多项式是最小相位的（即它的全部零点均在单位圆内），第二个则是最大相位的（全部零点均在单位圆外）。

将(12-1-13)式代入(12-1-12)式，得到

$$r_{xx}(k)=\frac{2B}{2\pi jT}\oint\frac{Z^{k-1}}{G_M(Z)G_M^*(1/Z^*)}dZ,\quad 0\leqslant k\leqslant M \tag{12-1-15}$$

利用(12-1-15)及(12-1-14)式，并取$T=1/2B$（T为采样周期，它满足乃奎斯特采样定理的要求），可以形成和式

$$\begin{aligned}\sum_{n=0}^{M}g_nr_{xx}(k-n)&=\frac{4B^2}{2\pi j}\oint\left\{\left[Z^{k-1}\sum_{n=0}^{M}g_nZ^{-n}\right]\Big/\left[G_M(Z)G_M^*(1/Z^*)\right]\right\}dZ\\&=\frac{4B^2}{2\pi j}\oint\frac{Z^{k-1}}{G_M^*(1/Z^*)}dZ,\quad 0\leqslant k\leqslant M\end{aligned} \tag{12-1-16}$$

利用柯西留数定理，知上述积分当$1\leqslant k\leqslant M$时其值为零。但对$k=0$，被积函数有一个单极点$Z=0$，留数为$1/g_0^*$。因此可将(12-1-16)式写成

$$\sum_{n=0}^{M}g_nr_{xx}(k-n)=\begin{cases}0 & k=1,2,\cdots,M\\(1/g_0^*)4B^2 & k=0\end{cases} \tag{12-1-17}$$

将上式归一化到g_0，得到

$$\sum_{n=0}^{M}a_nr_{xx}(k-n)=\begin{cases}0 & k=1,2,\cdots,M\\P_M & k=0\end{cases} \tag{12-1-18}$$

其中$a_n=g_n/g_0$，$P_M=(4B^2)/(g_0^*g_0)$。(12-1-18)式称为最大熵方程。

由(12-1-10)、(12-1-13)及(12-1-14)式立刻可得到用系数a_n表示的谱估计的表达式

$$\hat{S}_x(f)=\frac{P_M}{2B\left|1+\sum_{n=1}^{M}a_n e^{-j2\pi fnT}\right|^2} \tag{12-1-19}$$

(12-1-19)即是最大熵谱估值。系数a_n满足(12-1-18)式所到的$M+1$个最大熵方程。后面的分析将表明最大熵谱与自回归模型是等价的。

§12-2 预测误差滤波器(PEF)

使一个平稳过程的熵最大并不是唯一的得出上节(12-1-18)式的方程和谱(12-1-19)的方法。利用自化滤波器的概念，可从完全不同的角度得出同样的结果，亦即从维纳(Wiener)滤波的基本理论出发来推导出上述方程。

设滤波器是一个横向滤波器，其参数为

$$h_n: h_0,\ h_1,\ \cdots, h_{M-1} \tag{12-2-1}$$

即脉冲响应函数为h_n，因而滤波器的输出是输入序列$x(n)$与滤波器系数的卷积

$$\hat{z}_n=\sum_{k=0}^{M-1}h_k x_{n-k} \tag{12-2-2}$$

误差功率为

$$\begin{aligned}P&=E[(z_n-\hat{z}_n)(z_n-\hat{z}_n)^*]\\&=E\left[(z_n-\sum_{k=0}^{M-1}h_k x_{n-k})(z_n^*-\sum_{l=0}^{M-1}h_l^* x^*_{n-l})\right]\end{aligned} \tag{12-2-3}$$

其中z_n为期望的信号，P为实的正标量。可利用最小均方(LMS)准则，求得使P最小的一组h_n值。注意到(12-2-3)式中的滤波器的系数为复数，记为

$$\begin{aligned}h_m&=R_m+jI_m\\h_m^*&=R_m-jI_m\end{aligned} \tag{12-2-4}$$

使P达到极小值的必要条件为

$$\frac{\partial P}{\partial R_m}=\frac{\partial P}{\partial I_m}=0,\quad m=0,1,2,\cdots,M-1 \tag{12-2-5}$$

经过适当的运算不难得知，滤波器系数必须满足下列方程

$$\sum_{k=0}^{M-1}h_k r_{xx}(m-k)=r_{zx}(m),\quad m=0,1,\cdots,M-1 \tag{12-2-6}$$

其中$r_{zx}(m)$为z_n及x_n的互相关函数。上式即为维纳滤波方程，写成矩阵形式为

$$\begin{pmatrix}r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(1-M)\\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(2-M)\\ \cdots & \cdots & \cdots & \cdots\\ r_{xx}(M-1) & r_{xx}(M-2) & \cdots & r_{xx}(0)\end{pmatrix}\begin{pmatrix}h_0\\ h_1\\ \vdots\\ h_{M-1}\end{pmatrix}=\begin{pmatrix}r_{zx}(0)\\ r_{zx}(1)\\ \vdots\\ r_{zx}(M-1)\end{pmatrix} \tag{12-2-7}$$

对于一步预测的情况，期望输出为$z_n=x_{n+1}$，此时输入与输出序列的互相关函数为

$$\begin{aligned}r_{zx}(m)&=E[z_k x_{k-m}^*]=E[x_{k+1}x_{k-m}^*]\\&=E[x_k x_{k-m-1}^*]=r_{xx}(m+1)\end{aligned}$$

因此，滤波方程变为

$$\begin{pmatrix} r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(1-M) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(2-M) \\ \cdots & \cdots & \cdots & \cdots \\ r_{xx}(M-1) & r_{xx}(M-2) & \cdots & r_{xx}(0) \end{pmatrix}\begin{pmatrix} h_0 \\ h_1 \\ \vdots \\ h_{M-1} \end{pmatrix}=\begin{pmatrix} r_{xx}(1) \\ r_{xx}(2) \\ \vdots \\ r_{xx}(M) \end{pmatrix} \tag{12-2-8}$$

上式预测滤波方程常称为Yule-Walker方程。

由(12-2-3)式，利用(12-2-8)式不难得到预测滤波器的最小误差功率

$$P_{min}=r_{xx}(0)-\sum_{k=0}^{M-1}h_k r_{xx}(-k-1) \tag{12-2-9}$$

将(12-2-8)式右端移到左端，并利用(12-2-9)式，因而(12-2-8)式的矩阵增广为

$$\begin{pmatrix} r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(-M) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(-M+1) \\ \cdots & \cdots & \cdots & \cdots \\ r_{xx}(M) & r_{xx}(M-1) & \cdots & r_{xx}(0) \end{pmatrix}\begin{pmatrix} 1 \\ -h_0 \\ \vdots \\ -h_{M-1} \end{pmatrix}=\begin{pmatrix} P_M \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{12-2-10}$$

式中记$P_M=P_{min}$。由系数$1,-h_0,-h_1,\cdots,-h_{M-1}$定义的滤波器称为序列$x_n$的$M+1$阶预测误差滤波器（PEF）。容易看出，这一滤波器对序列x_n的响应是由下式定义的一步预测误差序列：

$$\varepsilon_{n+1}=x_{n+1}-\hat{x}_{n+1}=x_{n+1}-\sum_{k=0}^{M-1}h_k x_{n-k}$$

$$=\sum_{k=0}^{M}a_k x_{n-k+1} \tag{12-2-11}$$

其中系数a_k定义为

$$a_k=\begin{cases} 1 & k=0 \\ -h_{k-1} & k=1,2,\cdots,M \end{cases} \tag{12-2-12}$$

由(12-2-11)式很易得出预测误差滤波器的传输函数

$$A(Z)=\sum_{k=0}^{M}a_k Z^{-k}=1+\sum_{k=1}^{M}a_k Z^{-k} \tag{12-2-13}$$

预测误差滤波器的输出ε_{n+1}是白噪声化序列，亦即它将输入端的所有可预测部分都去掉了，而产生了一个尽可能接近白噪声的输出(在最小均方意义下)。

设 PEF 输出是精确的白噪声序列，其功率谱密度为$P_M/2B$，其输入序列x_n的谱可由PEF的传输函数(12-2-13)求得如下

$$\hat{S}_x(f)=\frac{P_M/2B}{\left|1+\sum_{k=1}^{M}a_k e^{-j2\pi fkT}\right|^2} \tag{12-2-14}$$

可以看出(12-2-10)、(12-2-14)式和(12-1-18)、(12-1-19)式完全相同。

§12-3 自回归模型法与最大熵谱分析法的等价

以上两节导出了最大熵谱，并说明了最大熵谱与维纳滤波一步预测所得到的谱是一致的。这一节我们将证明最大熵方法与自回归模型是等价的，它们之间存在着密切的关系，

正是揭示了这种关系，才使早已发展的自回归模型在谱估计方面重新得到人们广泛的注意。Van Pen Bos[2]证明了最大熵方法对相关函数的外推与AR模型的外推是相同的，从而在本质上建立了二者的内在联系。

由(12-1-4b)式可知，一个离散平稳过程$x_n(n=1,2,\cdots,N)$的熵为

$$H=\frac{1}{2}\ln\{\det[R_x]\}$$

其中R_x为半正定的Toeplitz自相关矩阵。在已知头$M+1$个自相关函数值$r_{xx}(0),r_{xx}(1),\cdots,r_{xx}(M)$的情况下，按最大熵原则求$r_{xx}(M+1)$，即按必要条件$dH/dr_{xx}(M+1)=0$解出$r_{xx}(M+1)$。相关矩阵可写为

$$R_x(M+1)=\begin{pmatrix} r_{xx}(0) & r_{xx}(1) & \cdots & r_{xx}(M) & r_{xx}(M+1)\\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(M-1) & r_{xx}(M)\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ r_{xx}(M+1) & r_{xx}(M) & \cdots & r_{xx}(1) & r_{xx}(0)\end{pmatrix} \tag{12-3-1}$$

使H最大的必要条件等价于使下式成立

$$\det\begin{pmatrix} r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(M-1)\\ r_{xx}(2) & r_{xx}(1) & \cdots & r_{xx}(M-2)\\ \cdots & \cdots & \cdots & \cdots\\ r_{xx}(M+1) & r_{xx}(M) & \cdots & r_{xx}(1)\end{pmatrix}=0 \tag{12-3-2}$$

而对于M阶的AR过程有

$$x_n=a_1x_{n-1}+a_2x_{n-2}+\cdots+a_Mx_{n-M}+a_n \tag{12-3-3}$$

其中a_n为估计的误差。用x_{n-k}乘上式两端，取期望值，并注意到$E[a_ia_j]=0$（当$i\neq j$）以及$E[x_{n-k},\ a_n]=0$（当$K>0$），我们得到

$$r_{xx}(K)=a_1r_{xx}(K-1)+a_2r_{xx}(K-2)+\cdots+a_Mr_{xx}(K-M),\quad K>0 \tag{12-3-4}$$

将$K=1,2,\cdots,M+1$代入(12-3-4)式，得到一组方程

$$r_{xx}(1)-a_1r_{xx}(0)-a_2r_{xx}(1)-\cdots-a_Mr_{xx}(M-1)=0$$

$$r_{xx}(2)-a_1r_{xx}(1)-a_2r_{xx}(0)-\cdots-a_Mr_{xx}(M-2)=0$$

$$\cdots\cdots$$

$$r_{xx}(M+1)-a_1r_{xx}(M)-a_2r_{xx}(M-1)-\cdots-a_Mr_{xx}(1)=0 \tag{12-3-5}$$

在假定头$M+1$个延迟值的自相关函数$r_{xx}(K)$（$K=0,1,\cdots,M$）已知的情况下，可由(12-3-5)式的头M个方程解得$a_1,a_2\cdots,a_M$。因而未知值$r_{xx}(M+1)$可通过解下列方程而获得

$$\det\begin{pmatrix} r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(M-1)\\ r_{xx}(2) & r_{xx}(1) & \cdots & r_{xx}(M-2)\\ \cdots & \cdots & \cdots & \cdots\\ r_{xx}(M+1) & \cdots & r_{xx}(M) & r_{xx}(1)\end{pmatrix}=0 \tag{12-3-6}$$

比较(12-3-2)式和(12-3-6)式可知AR过程，与最大熵外推所得的$r_{xx}(M+1)$值完全相同。因而最大熵谱分析等价于用AR模型来拟合这一随机过程。

将(12-3-3)式的AR过程取Z变换得到

$$X(Z)-X(Z)(a_1Z^{-1}+a_2Z^{-2}+\cdots+a_MZ^{-M})=A(Z)$$

因而

$$|X(Z)|^2 = \frac{|A(Z)|^2}{\left|1 - \sum_{k=1}^{M} a_k Z^{-k}\right|^2} \tag{12-3-7}$$

将$Z = \exp(j2\pi f)$代入(12-3-7)式之后，得到与§12-1中（12-1-19）式类似的功率谱表示式。从而证明了AR谱与最大熵的等价。

§12-4　计算最大熵谱(或AR谱)的几种方法

由上面的分析可知，计算最大熵谱的问题集中在如何求(12-1-19)式的系数a_m。在得到a_m之后，便可得到谱估值。这一节将介绍几种常用的算法。

一、莱文森(Leyinson)算法

最大熵谱分析的基本思想是在已知过程t_n的头$M+1$个自相关函数值时，对未知的自相关值不作任何假设，按照使过程的熵最大的原则来估计过程的功率谱。莱文森算法就是在这一假设下利用一组递推公式求出PEF的系数，而不是直接去解最大熵方程(12-1-18)或Yule-Walker方程(12-2-8)。

假定过程x_n的头$M+1$个自相关函数值为$r_{xx}(0), r_{xx}(1), \cdots, r_{xx}(M)$。为了得到过程$x_n$的最大熵谱，必须知道预测误差滤波器(PEF)的功率输出P_M及该滤波器的系数a_m($m=1, 2 \cdots, M$)换言之，要解下列$M+1$个联立方程(见(12-1-18)式)

$$\begin{pmatrix} r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(-M) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(-M+1) \\ \cdots & \cdots & \cdots & \cdots \\ r_{xx}(M) & r_{xx}(M-1) & \cdots & r_{xx}(0) \end{pmatrix} \begin{pmatrix} 1 \\ a_1 \\ \vdots \\ a_M \end{pmatrix} = \begin{pmatrix} P_M \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{12-4-1}$$

假定我们已经完成了第m步运算，并得到了系数$a_{m,i}$($i=0, 1, \cdots, m$)这里下标i表示系数号。因此有

$$\begin{pmatrix} r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(-m) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(-m+1) \\ \cdots & \cdots & \cdots & \cdots \\ r_{xx}(m) & r_{xx}(m-1) & \cdots & r_{xx}(0) \end{pmatrix} \begin{pmatrix} a_{m,0} \\ a_{m,1} \\ \vdots \\ a_{m,m} \end{pmatrix} = \begin{pmatrix} P_m \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{12-4-2}$$

定义一个量

$$B_m = a_{m,0} r_{xx}(m+1) + a_{m,1} r_{xx}(m) + \cdots + a_{m,m} r_{xx}(1) \tag{12-4-3}$$

并增广（12-4-2）的矩阵为

$$\begin{pmatrix} r_{xx}(0) & r_{xx}(1) & \cdots & r_{xx}(-m) & r_{xx}(-m-1) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(1-m) & r_{xx}(-m) \\ \cdots & \cdots & \cdots & \cdots & \\ r_{xx}(m) & r_{xx}(m+1) & \cdots & r_{xx}(0) & r_{xx}(-1) \\ r_{xx}(m+1) & r_{xx}(m) & \cdots & r_{xx}(1) & r_{xx}(0) \end{pmatrix} \begin{pmatrix} a_{m,0} \\ a_{m,1} \\ \vdots \\ a_{m,m} \\ 0 \end{pmatrix} = \begin{pmatrix} P_m \\ 0 \\ \vdots \\ 0 \\ B_m \end{pmatrix} \tag{12-4-4}$$

（12-4-4）式可简记为

$$R_x(m+1) \begin{pmatrix} a_{m,0} \\ a_{m,1} \\ \vdots \\ a_{m,m} \\ 0 \end{pmatrix} = \begin{pmatrix} P_m \\ 0 \\ \vdots \\ 0 \\ B_m \end{pmatrix} \tag{12-4-5}$$

其中$R_x(m+1)$已知。(12-4-5)式两边各进行复共轭转置之后有

$$[a^*_{m,0}, a^*_{m,1}, \cdots, a^*_{m,m}, 0]R^H_x(m+1) = [P^*_m, 0, \cdots, 0, B^*_m] \qquad (12\text{-}4\text{-}6)$$

由于自相关矩阵$R(m+1)$是埃尔密特矩阵，故有

$$R^H_x(m+1) = R_x(m+1) \qquad (12\text{-}4\text{-}7)$$

因此(12-4-6)式可写为

$$[a^*_{m,0}, a^*_{m,1}, \cdots, a^*_{m,m}, 0]R_x{}^H(m+1) = [P^*_m, 0, \cdots, 0, B^*_m] \qquad (12\text{-}4\text{-}8)$$

又因$R_x(m+1)$是托布里兹(Taeplitz)矩阵，(12-4-8)式再转置后不难证明可写为

$$R_x(m+1)\begin{pmatrix} 0 \\ a^*_{m,m} \\ \vdots \\ a^*_{m,1} \\ a^*_{m,0} \end{pmatrix} = \begin{pmatrix} B^*_m \\ 0 \\ \vdots \\ 0 \\ P^*_m \end{pmatrix} \qquad (12\text{-}4\text{-}9)$$

(12-4-9)式两边乘以K_m与(12-4-5)式相加(其中K_m是待求常系数)得到

$$R_x(m+1)\begin{pmatrix} a_{m,0} \\ a_{m,1} + K_m a^*_{m,m} \\ \vdots \\ a_{m,m} + K_m a^*_{m,1} \\ K_m a^*_{m,0} \end{pmatrix} = \begin{pmatrix} P_m + K_m B^*_m \\ 0 \\ \vdots \\ 0 \\ B_m + K_m P^*_m \end{pmatrix} \qquad (12\text{-}4\text{-}10)$$

另一方面，在第$m+1$步，所要求的PEF系数满足$m+1$阶预测滤波方程

$$R_x(m+1)\begin{pmatrix} a_{m+1,0} \\ a_{m+1,1} \\ \vdots \\ a_{m+1,m+1} \end{pmatrix} = \begin{pmatrix} P_{m+1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad (12\text{-}4\text{-}11)$$

比较(12-4-10)和(12-4-11)式得到

$$\begin{aligned}
&a_{m+1,0} = a_{m,0} \\
&a_{m+1,1} = a_{m,1} + K_m a^*_{m,m} \\
&\cdots\cdots\cdots\cdots \\
&a_{m+1,m} = a_{m,m} + K_m a^*_{m,1} \\
&P_{m+1} = P_m + K_m B^*_m \\
&B_m + K_m P^*_m = 0 \\
&a_{m+1,m+1} = K_m a^*_{m,0}
\end{aligned} \qquad (12\text{-}4\text{-}12)$$

由上式递推公式知道，从第 m 步到第 $m+1$ 步，零阶系数保持不变。因为 $a_{m,0}=1$，故 $a_{m+1,0}=1$，从而有

$$a_{m+1,m+1} = K_m,\quad m = 0, 1, \cdots, M-1 \qquad (12\text{-}4\text{-}13)$$

以及$P_0 = r_{xx}(0)$，因此(12-4-12)式可写成

$$\begin{aligned}
&a_{m+1,1} = a_{m,1} + a_{m+1,m+1} a^*_{m,m} \\
&a_{m+1,2} = a_{m,2} + a_{m+1,m+1} a^*_{m,m-1} \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
&a_{m+1,m} = a_{m,m} + a_{m+1,m+1} a^*_{m,1}
\end{aligned} \qquad (12\text{-}4\text{-}14)$$

以及

$$P_{m+1} = P_m + a_{m+1,m+1} B^*_m \qquad (12\text{-}4\text{-}15)$$

$$B_m + a_{m+1,m+1}P_m^* = 0 \tag{12-4-16}$$

其中

$$B_m = r_{xx}(m+1) + a_{m,1}r_{xx}(m) + \cdots + a_{m,m}r_{xx}(1), \quad m = 0,1,\cdots,M \tag{12-4-17}$$

由(12-4-15)及(12-4-16)式容易得到

$$P_{m+1} = P_m(1 - |a_{m+1,m+1}|^2) \tag{12-4-18}$$

初始条件为$P_0 = r_{xx}(0)$。可以证明[6]，一个系数由递归方程(12-4-14)给出的线性非递归滤波器，当且仅当系数 $a_{i,i}$ 满足条件

$$|a_{i,i}| < 1 \qquad i = 1,2,\cdots,M \tag{12-4-19}$$

时是最小相传的。但若(12-4-1)式左边是一相关矩阵(赫尔密特、托布里兹以及正定)时，(12-4-19)式的条件总是满足的。Berkhout[7]证明了这点。

由(12-4-19) 式所示的特性可知，$|P_{m+1}| \leqslant |P_m|$，即随着PEF 阶数的增高，预测误差功率将减小或不变。$a_{i,i}$常称为反射系数，物理上它相当于二端纲络终端不匹配时的反射程度。

(12-4-14)～(12-4-17) 式为确定预测误差滤波器参数的递归公式。这一方法避免了直接进行矩阵求逆去解方程(12-4-1)。

三、伯格(Burg)算法

莱文森递归公式求PEF系数的方法要求知道头$M+1$个自相关函数值，然后逐步求得反射系数$a_{i,i}$以及PEF系数。但在实际上，往往只知道时间序列x_n的有限数据，必须只利用x_n的有效数据对相关值进行估计。常规方法(第十一章中估计相关函数的方法)，或者假定有效数据区间外的数据为零，或者假定是周期扩展，这两种假设都是不合适的。伯格[14][15]提出的方法对有效数据不作任何假设，而是利用了莱文森算法和PEF的特性去一步一步估计系数a_m，同时又估计了相关值 $r_{xx}(m)$。(12-4-1)式表明了头$M+1$个相关值与$M+1$点预测误差滤波器及其功率输出之间的关系。它们之中的任一个都可用来估计其余二者。因此可以首先估计相应的预测误差滤波器的系数来估计$M+1$ 个相关值。这就是伯格算法的主要思想。

首先从$M=0$开始，(12-4-1)式的方程退化为

$$r_{xx}=(0)=P_0$$

然后假定零延迟的自相关函数

$$r_{xx}(0) = \frac{1}{N}\sum_{j=1}^{N} x_j x_j^* \tag{12-4-20}$$

为有效数据的平均功率。再考虑$M=1$的情况。将(12-4-1)式写为可与$M=0$时的解相对照的形式

$$\begin{pmatrix} P_0 & r_{xx}(-1) \\ r_{xx}(1) & P_0 \end{pmatrix}\begin{pmatrix} 1 \\ a_{1,1} \end{pmatrix}=\begin{pmatrix} P_1 \\ 0 \end{pmatrix} \tag{12-4-21}$$

上式有三个变量：$r_{xx}(1)$、$a_{1,1}$、P_1，因而需要有另一个方程。故采用假设：反射系数$a_{1,1}$必须选择得使这个两点 PEF 对有效数据工作得最好，亦即它的输出功率最小。前向两点和后向两点PEF的输出为

$$\begin{aligned} \hat{e}_{1,j} &= x_j + a_{1,1}x_{j-1} \\ \check{e}_{1,j} &= x_{j-1} + a_{1,1}^* x_j \end{aligned} \tag{12-4-22}$$

且有

$$\hat{e}_{0,j}=\check{e}_{0,j}=x_j$$

不难看出，对于一个平稳时间序列，前向和后向 PEF 的输出功率相等，亦即它们对有效数据工作得一样好。但是问题是，利用使前向或后向 PEF 输出功率最小化来确定反射系数，可能导致反射系数幅值大于 1，即不能保证 PEF 的稳定性。因此采用使两者平均值最小的方法来求反射系数。平均误差功率为

$$P_1=\frac{1}{2(N-1)}\sum_{j=1}^{N-1}\{|x_{j+1}+a_{1,1}x_j|^2+|x_j+a_{1,1}^*x_{j+1}|^2\} \tag{12-4-23}$$

使P相对于$a_{1,1}$最小，亦即使$\partial P/\partial a_{1,1}=0$，故得到

$$a_{1,1}=-\frac{2\sum_{j=1}^{N-1}x_{j+1}x_j^*}{\sum_{j=1}^{N-1}(x_{j+1}x_{j+1}^*+x_jx_j^*)}=-\frac{2\sum_{j=1}^{N-1}\hat{e}_{0,j+1}\check{e}_{0,j}}{\sum_{j=1}^{N-1}(|\hat{e}_{0,j+1}|^2+|\check{e}_{0,j}|^2)} \tag{12-4-24}$$

容易证明有 $|a_{1,1}|\leqslant 1$。由(12-4-21)式可解得$r_{xx}(1)$和P_1如下

$$r_{xx}(1)=r_{xx}^*(-1)=-a_{1,1}r_{xx}(0)$$
$$P_1=(1-a_{1,1}a_{1,1}^*)P_0 \tag{12-4-25}$$

下一步考虑三点滤波的情况。为确保满足莱文森算法，必须有下列关系成立(见(12-4-14)、(12-4-15)及(12-4-16)式)

$$P_2=P_1+a_{2,2}\Delta_1^*$$
$$\Delta_1+a_{2,2}P_1^*=0 \tag{12-4-26}$$
$$a_{2,1}=a_{1,1}+a_{2,2}a_{1,1}^*$$

因而(12-4-1)式变为

$$\begin{pmatrix} r_{xx}(0) & r_{xx}(-1) & r_{xx}(-2)\\ r_{xx}(1) & r_{xx}(0) & r_{xx}(-1)\\ r_{xx}(2) & r_{xx}(1) & r_{xx}(0)\end{pmatrix}\left(\begin{vmatrix}1\\ a_{1,1}\\ 0\end{vmatrix}+a_{2,2}\begin{vmatrix}0\\ a_{1,1}^*\\ 1\end{vmatrix}\right)=\left(\begin{vmatrix}P_1\\ 0\\ \Delta_1\end{vmatrix}+a_{2,2}\begin{vmatrix}\Delta_1^*\\ 0\\ P_1^*\end{vmatrix}\right) \tag{12-4-27}$$

此时平均功率输出为

$$P=\frac{1}{2(N-2)}\sum_{j=1}^{N-2}[|x_{j+2}+a_{2,1}x_{j+1}+a_{2,2}x_j|^2+|x_j+a_{2,1}^*x_{j+1}+a_{2,2}^*x_{j+2}|^2] \tag{12-4-28}$$

利用使P最小的条件$\partial P/\partial a_{2,2}=0$，可以解得

$$a_{2,2}=\frac{-2\sum_{j=1}^{N-2}(x_{j+2}+a_{1,1}x_{j+1})(x_j+a_{1,1}^*x_{j+1})^*}{\sum_{j=1}^{N-2}(|x_{j+2}+a_{1,1}x_{j+1}|^2+|x_j+a_{1,1}^*x_{j+1}|^2)}$$

$$=-\frac{2\sum_{j=1}^{N-2}\hat{e}_{1,j+2}\check{e}_{1,j+1}}{\sum_{j=1}^{N-2}(|\hat{e}_{1,j+2}|^2+|\check{e}_{1,j+1}|^2)} \tag{12-4-29}$$

由(12-4-27)式可解得$r_{xx}(2)$、P_2如下

$$r_{xx}(2) = -a_{2,1} r_{xx}(1) - a_{2,2} r_{xx}(0)$$

$$P_2 = (1 - a_{2,2} a^*_{2,2}) P_1$$

其中

$$a_{2,1} = a_{1,1} + a_{2,2} a^*_{1,1}$$

继续进行这一过程，即写出形如(12-4-1)式的正规方程和满足莱文森解的关系式，然后按照使 PEF 平均输出功率最小的条件求出反射系数$a_{i,i}$，最后解正规方程得到下一组新的$r_{xx}(i)$以及P_i。在M次递推之后可得到系数$(1, a_{M,1}, a_{M,2}, \cdots, a_{M,M})$如下

$$a_{M,M} = \frac{-2 \sum_{j=1}^{N-M} \hat{e}_{M-1,j+M} \check{e}^*_{M-1,j+M-1}}{\sum_{j=1}^{N-M} \left(|\hat{e}_{M-1,j+M}|^2 + |\check{e}_{M-1,j+M-1}|^2 \right)}$$

$$a_{M,1} = a_{M-1,1} + a_{M,M} a^*_{M-1,M-1}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$a_{M,M-1} = a_{M-1,M-1} + a_{M,M} a^*_{M-1,1} \tag{12-4-30a}$$

以及

$$P_M = P_{M-1}(1 - a_{M,M} a^*_{M,M})$$

$$r_{xx}(M) = -\sum_{j=1}^{M} a_{M,j} r_{xx}(M-j) \tag{12-4-30b}$$

其中

$$\hat{e}_{M,j} = \sum_{m=0}^{M} a_{M,m} x_{j-m} = \hat{e}_{M-1,j} + a_{M,M} \check{e}_{M-1,j-1}$$

$$\check{e}_{M,j} = \sum_{m=0}^{M} a^*_{M,M-m} x_{j-m} = \check{e}_{M-1,j-1} + a^*_{M,M} \hat{e}_{M-1,j} \tag{12-4-31}$$

(12-4-31)式很易求得，只要将莱文森递推公式(12-4-14)代入$\hat{e}_{M,j}$表达式即可。很明显，伯格算法导致了一个满足条件$|a_{i,i}| \leqslant 1 (i = 1, 2, \cdots, M)$的 PEF。如前所述，因此导出的 PEF 是最小相位的，而且估计的自相关矩阵是埃尔密特、托布里兹和正定矩阵。倘若其中一个反射系数，例如$a_{M,M}$变为1，则由(12-4-18)式可知误差功率P_M变为零。这意味着对应的PEF 对时间序列的下一点进行了完全的预测。此时PEF 不是最小相位的，但其全部极点均在单位圆上，输入谱在频域上由一系列 δ-函数组成，伯格曾对此作了证明。

由(12-4-31)式很容易得出如图12-4-1所示的PEF的格型结构流图。

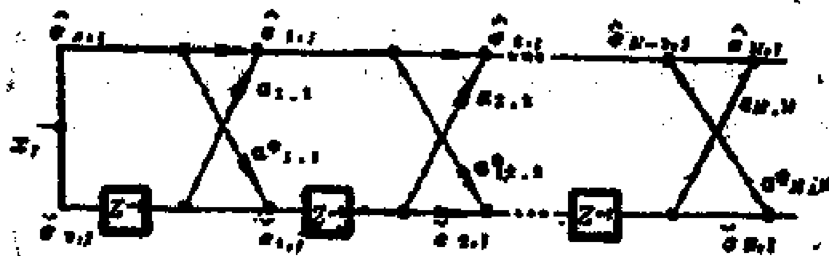


图12-4-1 PEF格型结构流图

三、最小二乘算法

伯格算法利用了莱文森算法的递推特性以及使前、后向预测误差平均功率对单个系数 $a_{i,i}(i=1,2,\cdots,M)$ 最小的方法，估计了PEF的系数及自相关值，最小二乘算法则是使预测误差功率对特定的PEF的全部系数最小来估计PEF的系数。

给定一平稳时间序列 $x_n(n=1,2,\cdots,N)$，前向预测值为 $\hat{x}_t$，期望值为 x_t，则预测误差滤波器的输出为

$$\hat{e}_t = x_t - \hat{x}_t = x_t - \sum_{k=1}^{M} a_k x_{t-k}, \qquad t = M+1,\cdots,N \tag{12-4-32}$$

类似地，后向PEF的输出为

$$\check{e}_t = x_t - \sum_{k=1}^{M} a^*_k x_{t+k}, \qquad t = 1,2,\cdots,N-M \tag{12-4-33}$$

(12-4-32)和(12-4-33)式可改写成

$$\hat{e}_t = \sum_{k=0}^{M} r_k x_{t-k}, \qquad t = M+1,\cdots,N \tag{12-4-32a}$$

$$\check{e}_t = \sum_{k=0}^{M} r_k^* x_{t+k}, \qquad t = 1,2,\cdots,N-M \tag{12-4-33a}$$

前向和后向预测误差功率输出分别为

$$\hat{P}_M = \sum_{t=M+1}^{N} \left[\left(\sum_{k=0}^{M} r_k x_{t-k} \right) \left(\sum_{l=0}^{M} r_l x_{t-l} \right)^* \right] \tag{12-4-34}$$

$$\check{P}_M = \sum_{t=1}^{N-M} \left[\left(\sum_{k=0}^{M} r_k^* x_{t+k} \right)^* \left(\sum_{l=0}^{M} r^*_l x_{t+l} \right) \right]$$

$$= \sum_{t=M+1}^{N} \left[\left(\sum_{k=0}^{M} r_k^* x_{t-M+k} \right)^* \left(\sum_{l=0}^{M} r_l^* x_{t-M+l} \right) \right] \tag{12-4-35}$$

使前向后向总功率最小的必要条件是 $\frac{\partial P_M}{\partial r_i^*} = 0\ (i=1,2,\cdots,M)$，这里 $P_M = \hat{P}_M + \check{P}_M$。利用(12-4-34)及(12-4-35)式和这个条件，并注意交换求和次序不难得到

$$\sum_{k=0}^{M} r_k \left[\sum_{t=M+1}^{N} x_{t-i}^* x_{t-k} + \sum_{t=M+1}^{N} x_{t-M+i} x_{t-M+k}^* \right] = 0 \qquad i = 1,2,\cdots,M \tag{12-4-36}$$

记

$$\sum_{k=0}^{M} r_k \left[\sum_{t=M+1}^{N} x_t^* x_{t-k} + \sum_{t=M+1}^{N} x_{t-M} x_{t-M+k}^* \right] = P \tag{12-4-37}$$

利用(12-4-37)式，将(12-4-36)式扩展之后得到 $M+1$ 个关于 $r_1, r_2, \cdots, r_M, P$ 的联立方程。用矩阵形式表示，可写为

$$\Phi r = p \tag{12-4-38}$$

其中

$$\Phi = [\varphi_{ij}] = \left[\sum_{t=M+1}^{N} x_{t-i+1}^* x_{t-j+1} + \sum_{t=M+1}^{N} x_{t-M+i-1} x_{t-M+j-1}^* \right] \tag{12-4-39a}$$

$$\mathbf{r}=[1,r_1,r_2,\cdots,r_M]^T \tag{12-4-39b}$$

$$\mathbf{p}=[P,0,\cdots,0]^T \tag{12-4-39c}$$

由(12-4-39a)表示的矩阵称为最小二乘矩阵。如该矩阵元素$\varphi_{i,j}$可以求出，则系数$\mathbf{r}$即可由解(12-4-36)式得到。因为

$$\begin{aligned}
\varphi_{i-1,j-1} &= \sum_{t=M+1}^{N} x^*_{t-i+2}\, x_{t-j+2} + \sum_{t=M+1}^{N} x_{t-M+i-2}\, x^*_{t-M+j-2} \\
&= \sum_{t=M+1}^{N} x^*_{t-i+1}\, x_{t-j+1} - x^*_{M-i+2}\, x_{M-j+2} + x^*_{N-i+2}\, x_{N-j+2} \\
&\quad + \sum_{t=M+1}^{N} x_{t-M+i-1}\, x^*_{t-M+j-1} + x_{i-1}\, x^*_{j-1} - x_{N-M+i-1}\, x^*_{N-M+j-1}
\end{aligned}$$

$$2\leqslant i,j\leqslant M+1 \tag{12-4-40}$$

比较$\varphi_{i-1,j-1}$与$\varphi_{i,j}$的表达式可以得到从$\varphi_{i-1,j-1}$求$\varphi_{i,j}$的递推公式为

$$\begin{aligned}
\varphi_{i,j} = \varphi_{i-1,j-1} &+ x^*_{M-i+2}\, x_{M-j+2} - x^*_{N-i+2}\, x_{N-j+2} \\
&+ x_{N-M+i-1}\, x^*_{N-M+j-1} - x_{i-1}\, x^*_{j-1}
\end{aligned} \tag{12-4-41}$$

利用

$$\varphi_{i,1} = \sum_{t=M+1}^{N} x^*_{t-i+1}\, x_t + \sum_{t=M+1}^{N} x_{t-M+i-1}\, x^*_{t-M} \qquad i=1,2,\cdots,M+1$$

作为初始值。对于一个给定的阶数M，先求最小二乘矩阵Φ的第一列，然后利用递推公式(12-4-41)式求$\varphi_{i,j}$。采用这种方法求最小二乘矩阵比直接按（12-4-39a）求这个矩阵，计算量可大大减小。乘法次数由$2NM^2-2M^3+0(M^2)$次削减到$2MN+0(M^2)$次。

四、改进的最小二乘算法

上面所述最小二乘算法由于形成的最小二乘矩阵并不具有相关矩阵特性，因而可能导致估计的PEF不是最小相位的。改进的算法直接从正规方程出发一步一步形成正规方程，可保证得到的正规矩阵是托布里兹矩阵，从而保证了逆矩阵的存在。

再一次分别考虑对序列$x_n(n=1,2,\cdots,N)$的前向和后向PEF 的总误差输出功率

$$\begin{aligned}
\widehat{P}_M &= \sum_{t=M+1}^{N} (x_t-\widehat{x}_t)(x_t-\widehat{x}_t)^* \\
&= \sum_{t=M+1}^{N} \left[\left(x_t-\sum_{k=1}^{M} a_k x_{t-k}\right)\left(x_t^*-\sum_{k=1}^{M} a_k^* x_{t-k}^*\right)\right]
\end{aligned} \tag{12-4-42}$$

按最小二乘原则，当

$$\frac{\partial \widehat{P}_M}{\partial a_i^*} = \sum_{t=M+1}^{N} (-x^*_{t-i})\left(x_t-\sum_{k=1}^{M} a_k x_{t-k}\right)=0 \quad i=1,2,\cdots,M \tag{12-4-43}$$

时，总的前向PEF的功率输出最小，写为矩阵形式有

$$\widehat{X}^H\widehat{e}=0 \tag{12-4-44}$$

或

$$\widehat{X}^H(\widehat{y}-\widehat{X}a)=0 \tag{12-4-45}$$

或

$$\hat{X}^H\hat{X}a=\hat{X}^H\hat{y} \tag{12-4-46}$$

其中

$$\hat{y}=[x_{M+1},x_{M+2},\cdots,x_N]^T \tag{12-4-47}$$

$$a=[a_1,a_2,\cdots,a_M]^T \tag{12-4-48}$$

$$\hat{X}=\begin{pmatrix} x_M & x_{M-1} & \cdots & x_1 \\ x_{M+1} & x_M & \cdots & x_2 \\ \cdots & \cdots & \cdots & \cdots \\ x_{N-1} & x_{N-2} & \cdots & x_{N-M} \end{pmatrix} \tag{12-4-49}$$

进一步，方程(12-4-46)可写为

$$\hat{R}a=\hat{s} \tag{12-4-50}$$

其中

$$\hat{R}=\hat{X}^H\hat{X} \tag{12-4-51}$$

$$\hat{s}=[\hat{s}_1,\hat{s}_2,\cdots,\hat{s}_M]^T=\hat{X}^H\hat{y} \tag{12-4-52}$$

在这些方程中“H”，表示复共轭转置“T”，表示转置。

类似地，对于后向预测有

$$\check{X}^T(\check{y}^*-\check{X}^*a)=0 \tag{12-4-53}$$

或

$$\check{X}^T\check{X}^*a=\check{X}^T\check{y}^* \tag{12-4-54}$$

其中

$$\check{y}=[x_1,x_2,\cdots,x_{N-M}]^T \tag{12-4-55}$$

$$\check{X}=\begin{pmatrix} x_2 & x_3 & \cdots & x_{M+1} \\ x_3 & x_4 & \cdots & x_{M+2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{N-M+1} & x_{N-M+2} & \cdots & x_N \end{pmatrix} \tag{12-4-56}$$

方程(12-4-54)可进一步写成

$$\check{R}a=\check{s} \tag{12-4-57}$$

其中

$$\check{R}=\check{X}^T\check{X}^*, \quad \check{s}=\check{X}^T\check{y}^* \tag{12-4-58}$$

显然，$\hat{X}^H\hat{X}$和$\check{X}^T\check{X}^*$均为埃尔密特矩阵。

前向和后向总功率最小的条件变成

$$\hat{X}^H(\hat{y}-\hat{X}a)+\check{X}^T(\check{y}^*-\check{X}^*a)=0 \tag{12-4-59}$$

或

$$(\hat{X}^H\hat{X}+\check{X}^T\check{X}^*)a=\hat{X}^H\hat{y}+\check{X}^T\check{y}^* \tag{12-4-60}$$

若采用记号

$$X=\begin{pmatrix} \hat{X} \\ \cdots \\ \check{X}^* \end{pmatrix} \qquad y=\begin{pmatrix} \hat{y} \\ \cdots \\ \check{y}^* \end{pmatrix} \tag{12-4-61}$$

(12-4-60)式可写为

$$R a = s \tag{12-4-62}$$

其中

$$R = X^H X$$

$$s = X^H y = [s_1, s_2, \cdots, s_M]^T \tag{12-4-63}$$

(12-4-62)是描述系数a的$M\times M$复正规方程。它的解是唯一的。

我们的任务是根据给定的一组时间序列来确定R与s的元素，然后解这个正规方程。为清楚地说明算法，从$M=2$的前向PEF开始，看看如何由$\hat{R}_2$和$\hat{s}_2$求得$\hat{R}_3$和$\hat{s}_3$。

当$M=2$时，有

$$\hat{R}_2 = \begin{pmatrix} \hat{r}_{11}^{(2)} & \hat{r}_{12}^{(2)} \\ \hat{r}_{21}^{(2)} & \hat{r}_{22}^{(2)} \end{pmatrix} = \begin{pmatrix} \sum_{l=1}^{N-2} x^*_{l+1} x_{l+1} & \cdot \\ \sum_{l=1}^{N-2} x_l^* x_{l+1} & \sum_{l=1}^{N-2} x_l^* x_l \end{pmatrix}$$

$$\hat{s}_2 = [\hat{s}_1^{(2)}, \hat{s}_2^{(2)}]^T = \left[\sum_{l=1}^{N-2} x^*_{l+1} x_{l+2}, \sum_{l=1}^{N-2} x_l^* x_{l+2}\right]^T$$

这里$\hat{R}_2$对角线上方元素被省略，因为$\hat{R}_2$是埃尔密特矩阵，有$\hat{r}_{i,j} = \hat{r}_{j,i}^*$。

而当$M=3$时有

$$\hat{R}_3 = \begin{pmatrix} \hat{r}_{11}^{(3)} & \cdot & \cdot \\ \hat{r}_{21}^{(3)} & \hat{r}_{22}^{(3)} & \cdot \\ \hat{r}_{31}^{(3)} & \hat{r}_{32}^{(3)} & \hat{r}_{33}^{(3)} \end{pmatrix} = \begin{pmatrix} \sum_{l=1}^{N-3} x^*_{l+2} x_{l+2} & \cdot & \cdot \\ \sum_{l=1}^{N-3} x^*_{l+1} x_{l+2} & \sum_{l=1}^{N-3} x^*_{l+1} x_{l+1} & \cdot \\ \sum_{l=1}^{N-3} x_l^* x_{l+2} & \sum_{l=1}^{N-3} x^*_l x_{l+1} & \sum_{l=1}^{N-3} x_l^* x_l \end{pmatrix}$$

$$\hat{s}_3 = [\hat{s}_1^{(3)}, \hat{s}_2^{(3)}, \hat{s}_3^{(3)}]^T = \left[\sum_{l=1}^{N-3} x^*_{l+2} x_{l+3}, \sum_{l=1}^{N-3} x^*_{l+1} x_{l+3}, \sum_{l=1}^{N-3} x_l^* x_{l+3}\right]^T$$

比较两组$\hat{R}$和$\hat{s}$有

$$\hat{r}_{i,j}^{(3)} = \hat{r}_{i,j}^{(2)} - x^*_{3-i} x_{3-j}, \quad 1 \leqslant j \leqslant i \leqslant 2$$

$\hat{R}_3$的最后一行的元素(除第一元素$\hat{r}_{3,1}$外)为

$$\hat{r}_{3,1}^{(3)} = \hat{s}_2^{(2)} - x^*_{N-2} x_N$$

$$\hat{r}_{3,j}^{(3)} = \hat{r}_{2,j-1}^{(2)} - x^*_{N-2} x_{N+1-j}, \quad 2 \leqslant j \leqslant 3$$

而

$$\hat{s}_i^{(3)} = \hat{s}_i^{(2)} - x^*_{3-i} x_3, \quad 1 \leqslant i \leqslant 2$$

以及

$$\hat{s}_3^{(3)} = \sum_{l=1}^{N-3} x_l^* x_{l+3}$$

类似的方法可用于后向预测，有

$$\check{r}_{i,j}^{(3)} = \check{r}_{i,j}^{(2)} - x_{N-2+i} x^*_{N-2+j}$$

$\check{R}_3$的最后一行元素(除第一元素之外)为

$$\check{r}_{3,j}^{(3)} = \check{r}_{2,j-1}^{(2)} - x_j x_3^*, \quad 2 \leqslant j \leqslant 3$$

以及

$$\check{r}_{3,1}^{(3)} = \check{s}_2^{(2)} - x_3 x_1^*$$

而$\check{S}_3$的元素是

$$\check{s}_i^{(3)} = \check{s}_i^{(2)} - x_{N-2+i} x^*_{N-2}, \quad 1 \leqslant i \leqslant 2$$

以及

$$s_3 = \sum_{l=1}^{N-3} x_{l+3} x_l^*$$

由上述过程可见，当由$M=2$推导$M=3$的正规方程时，只有一个内积$\hat{s}_3^{(3)}$（或$\check{s}_2^{(3)}$）要计算。$\hat{R}_3$（或$\check{R}_3$）和$\hat{S}_3$（或$\check{S}_3$）的其余元素的计算，只各需要一次复数乘法。

由于$\hat{s}_{M+1}^{(M+1)} = \check{s}_{M+1}^{(M+1)}$，故当由$R_2$和$S_2$计算$R_3$和$S_3$时，只需计算一次内积，$R_3$和$S_3$的其余元素可以由$R_2$和$S_2$用两次复数乘法得到。

如上所述，不难得到由R_M和S_M计算R_{M+1}和S_{M+1}的公式

$$\begin{cases} r_{M+1,1}^{(M+1)} = s_M^{(M)} - x^*_{N-M} x_N - x_1^* x_{M+1} \\ r_{M+1,j}^{(M+1)} = r_{M,j-1}^{(M)} - x^*_{N-M} x_{N+1-j} - x_j^* x_{M+1}, \quad 2 \leqslant j \leqslant M+1 \\ r_{i,j}^{(M+1)} = r_{i,j}^{(M)} - x^*_{M+1-i} x_{M+1-j} - x_{N-M+i} x^*_{N-M+j}, \quad 1 \leqslant i \leqslant M, \; 1 \leqslant j \leqslant i \end{cases} \tag{12-4-64}$$

$$\begin{cases} s_i^{(M+1)} = s_i^{(M)} - x^*_{M+1-i} x_{M+1} - x^*_{N-M} x_{N-M+i}, \quad 1 \leqslant i \leqslant M, \; 1 \leqslant j \leqslant i \\ s_{M+1}^{(M+1)} = 2 \sum\limits_{l=1}^{N-M-1} x_l^* x_{M+1+l} \end{cases} \tag{12-4-65}$$

从$M=1$开始，首先计算

$$r_{1,1}^{(1)} = \sum_{l=1}^{N-1} x_l^* x_l + \sum_{l=2}^{N} x_l^* x_l = 2\sum_{l=2}^{N-1} x_l^* x_l + x_1^* x_1 + x_N^* x_N \tag{12-4-66}$$

$$s_1^{(1)} = \sum_{l=1}^{N-1} x_l^* x_{l+1} + \sum_{l=1}^{N-1} x_{l+1} x_l^* = 2\sum_{l=1}^{N-1} x_l^* x_{l+1} \tag{12-4-67}$$

这一方法是在已知滤波器阶数M时进行的。尽管对于计算最大熵谱时阶数M是一比较难于选定的值，但这算法提供了用计算手段动态选择阶数的可能性。此外，这一算法直接形成正规方程，可保证滤波器的稳定性。

上述算法并未包括如何计算残留平方和（PEF 输出功率）的项，这一项可按下式单独计算

$$P_M = \hat{P}_M + \check{P}_M = \mathbf{y}^H \mathbf{y} - \mathbf{a}^H \mathbf{s} \tag{12-4-68}$$

作为一个实例，图12-4-2画出了用这一方法计算和用伯格法计算的结果。两单频信号频率分别为1Hz和1.2Hz，噪声为零均值高斯白噪声，采样频率为20Hz，信号长度为1.5周期（31点）两信号初相位均为零，阶数$M=9$。结果表明，在两个信号的情况，这一算法优于Burg算法。当$M=9$时Burg算法已有峰分裂，因而不能区分两个谱峰。M增大，分裂更为严重。

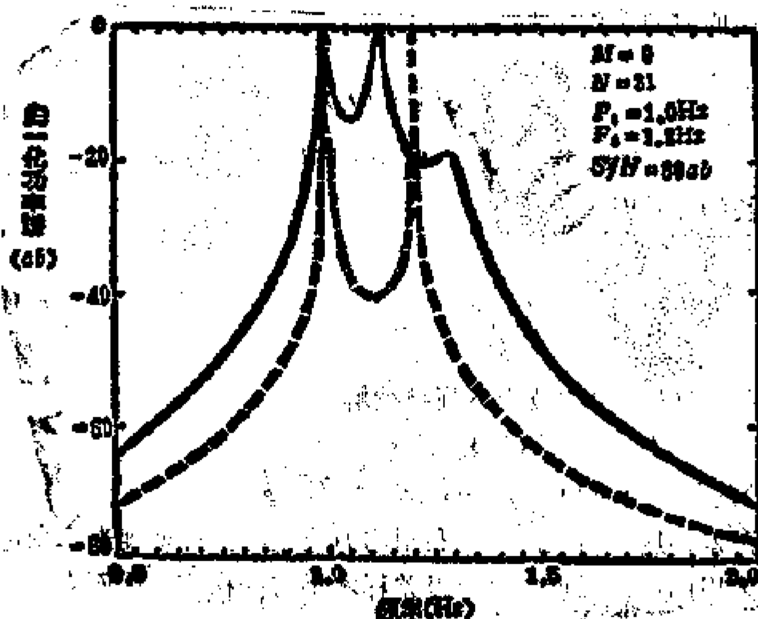


图12-4-2 伯格算法与最小二乘算法谱估计

五、自适应最小均方算法(自适应AR过程)

到目前为止、已经讨论了平稳时间序列的最大熵谱的算法。在非平稳过程的情况，采用自适应最小均方算法来求 PEF 的系数并用于谱分析，不仅方法简单，而且可能进行实时应用。

利用一个线性时不变预测滤波器的一步预测问题已在本章§ 12-2中讨论过。其最佳滤波器的系数是正规方程的解。这里假定线性预测滤波器是时变的，它的系数随着一个新数据样本的接收而被刷新。设预测滤波器在时刻k的输出(前向预测)为

$$\hat{x}(k)=\sum_{m=1}^{M}a_m(k)x(k-m),\quad k=M+1,\cdots,N$$

利用记号

$$A(k)=[a_1(k),\ a_2(k),\cdots,a_M(k)]^T \tag{12-4-69}$$

表示滤波器在时刻k的系数矢量。

预测误差为

$$\begin{aligned}e(k)&=x(k)-\hat{x}(k)\\&=x(k)-A^T(k)X(k-1)\end{aligned} \tag{12-4-70}$$

其中

$$X^T(k-1)=[x(k-1),x(k-2),\cdots,x(k-M)] \tag{12-4-71}$$

设$x(k)$为零均值平稳过程，误差平方的均值为

$$\begin{aligned}E[e^2(k)]=\overline{e^2(k)}&=E[(x(k)-A^T(k)X(k-1))(x(k)-X^T(k-1)A(k))]\\&=r_{xx}(0)-2P_x^TA(k)+A^T(k)R(M)A(k)\end{aligned} \tag{12-4-72}$$

其中

$$\begin{aligned}P_x&=E[x(k)X^T(k-1)]\\&=[E[x(k)x(k-1)],E[x(k)x(k-2)],\cdots,E[x(k)x(k-M)]]\end{aligned} \tag{12-4-73}$$

$R(M)$为序列的自相关矩阵，(12-4-72)式有唯一极小值，采用最速下降法，使滤波器系

数矢量沿误差表面负梯度方向迭代，即

$$A(k+1)=A(k)-\frac{\mu}{2}\nabla[E[e^2(k)]]$$

$$=A(k)-\mu(P_x-R(M)A(k)) \tag{12-4-74}$$

其中μ是调整迭代步长的标量因子。

可以证明，在

$$0<\mu<\frac{2}{\lambda_{max}} \tag{12-4-75}$$

的条件下，当$k\to\infty$时，$A(k+1)$收敛于下列维纳解[9]

$$A_{opt}=R^{-1}(M)P_x \tag{12-4-76}$$

其中λ_{max}为矩阵$R(M)$的最大特征值。

由于不知道均方误差梯度的精确值，故利用均方误差的估计梯度来实现最速下降法。自适应最小均方算法采用的估计梯度是单个误差样本平方的梯度。因而有

$$\hat{\nabla}_k=\nabla[e_k^2]=2e_k\nabla e_k \tag{12-4-77}$$

其中$\hat{\nabla}_k$是在时刻k的估计梯度。将(12-4-77)式代入(12-4-74)式得到

$$A(k+1)=A(k)-\mu e_k\nabla e_k \tag{12-4-78}$$

由(12-4-70)式，容易证明

$$\nabla e_k=-X(k-1) \tag{12-4-79}$$

因此

$$A(k+1)=A(k)+\mu e_k X(k-1)$$

或写为另一形式

$$A(k+1)=A(k)+\mu(x(\hat{k})-x(k))X(k-1) \tag{12-4-80}$$

由(12-4-80)式可知，下一个滤波器系数矢量可由现在的系数矢量加上一个输入矢量的修正量来获得。

由于

$$\lambda_{max}<t_{race}(R(M))=MR(0) \tag{12-4-81}$$

因而(12-4-75)式的收敛条件变为

$$0<\mu<\frac{2}{MR(0)} \tag{12-4-82}$$

或写为

$$\mu=\frac{a}{MR(0)},\quad 0<a<2 \tag{12-4-83}$$

其中$R(0)$是输入序列的功率，实际上它可用输入样本平方的平均值来近似。

收敛因子μ（或a）直接关系到算法的收敛速度。一般说，收敛时间常数是难于计算的参数，它与R的特征值有关。瞬态过程用时间常数来描述，在假定$R(M)$所有特征值相同时，时间常数表为

$$\tau=\frac{1}{\mu\lambda}=\frac{M}{a} \tag{12-4-84}$$

一般说，当ε较大时，收敛过程快，且有一个较大的稳态误差。反之，ε较小，收敛较快，且有一较小的稳态误差。

当样本序列为复数时，容易将(12-4-80)式推广为下列形式

$$A(k+1)=A(k)+\mu(x(k)-\hat{x}(k))X^*(k-1) \qquad (12\text{-}4\text{-}85)$$

初始条件可选为$A(0)=0$，这种选择只对头几次迭代有影响。

自适应最小均方算法的主要优点是简单，刷新系数所用的运算次数(乘法和加法)正比于系数的个数M。此外，它不必估计自相关值而最小二乘法除估计自相关值之外，尚需至少与M^2正比的运算次数。

§12-5 PEF(或AR模型)阶数的确定

在讨论最大熵谱时并未说明PEF的阶数如何确定。实际上最大熵谱的质量在很大程度上依赖于PEF的阶数。在低噪声情况下，阶数过高将会导致谱线分裂，阶数太低又得不到好的谱分辨力。

Akaike[10]提出了一个"客观"的准则，他对一被拟合的AR模型定义一个质量因素——最终预测误差(FPE)。它的含义是，当对序列x_n进行拟合的AR模型被用于另一序列y_n时，预计产生的预测误差。

设$\hat{x}_n$为时间序列x_n的最佳预测，滤波器阶数为M。其均方预测误差为

$$E[(x_n-\hat{x}_n)^2]=E\left[\left(x_n-\sum_{k=1}^{M}a_{x,k}x_{n-k}\right)^2\right] \qquad (12\text{-}5\text{-}1)$$

其中$a_{x,k}$为对x_n的最佳PEF的系数。考虑另一过程y_n，它与x_n相互独立，但具有同样的统计特性。把(12-5-1)式中的同一AR模型用于预测y_n，则有

$$\hat{y}_n=\sum_{k=1}^{M}a_{x,k}y_{n-k} \qquad (12\text{-}5\text{-}2)$$

均方误差为

$$\text{FPE}=E[(y_n-\hat{y}_n)^2] \qquad (12\text{-}5\text{-}3)$$

再用$a_{y,k}$表示由过程y_n按LMS准则所确定的AR模型的第k个系数，并记

$$\Delta a_{M,k}=a_{x,k}-a_{y,k} \qquad (12\text{-}5\text{-}4)$$

由(12-5-1)、(12-5-2)、(12-5-3)、(12-5-4)式容易得到

$$\text{FPE}=P_M+\sum_{m=1}^{M}\sum_{k=1}^{M}E[\Delta a_{M,m}\Delta a_{M,k}]E[y_{n-m}y_{n-k}] \qquad (12\text{-}5\text{-}5)$$

其中

$$P_M=E[(y_n-\sum_{m=1}^{M}a_{y,m}y_{n-m})^2] \qquad (12\text{-}5\text{-}6)$$

为将系数$a_{y,k}(k=1,\cdots,M)$的AR模型用于y_n时的最小均方误差。(12-5-6)式中的第一项随M增大而减小；相反，第二项将增大。这是由于对有限长的x_n、$a_{x,k}$和$a_{y,k}$有统计偏差所致。因而应当选取使FPE最小的M值。可以证明，当x_n均值为零时，对于阶数M的拟合，使FPE取最小估值的表示式为

$$(\mathrm{FPE})_M=\frac{N+M+1}{N-M-1}P_M \tag{12-5-7}$$

其中P_M为滤波器输出的误差功率。M为AR模型的阶数，N为数据点数。可以看出，P_M随M增大而减小，而$(N+M+1)/(N-M-1)$随M增大而增大。因而FPE_M将在某一$M=M_{opt}$处取最小值。

Akaike 后来又提出了另一个所谓信息论准则(AIC)[11]。他将最大似然原理推广应用于时间序列的统计假设检验，引进了AIC作为模型拟合的尺度。最终导出的AIC表示式为

$$\mathrm{AIC}_M=\ln(P_M)+\frac{2M}{N} \tag{12-5-8}$$

使AIC_M最小的M值作为最佳阶数M_{opt}。由(12-5-7)及(12-5-8)式容易得知，FPE_M与AIC_M是渐近等价的。亦即有

$$\lim_{N\to\infty}\{\ln \mathrm{FPE}_M\}=\mathrm{AIC}_M \tag{12-5-9}$$

有关两种准则关系的细节可参考文献[12]3.6节。

Parzen[13]提出了第三个准则，称为自回归传递函数准则(CAT)其计算公式为

$$\mathrm{CAT}_M=\frac{1}{N}\sum_{m=1}^{K}\frac{N-m}{NP_m}-\frac{N-M}{NP_M} \tag{12-5-10}$$

上述三个准则都可在设计 PEF 的算法的每一次迭代过程中进行计算。当某一准则新的计算结果比上一次迭代所得值大时，迭代终止，从而得到最佳阶数M_{opt}。但一般说FPE准则优于其它两种准则。大量计算结果表明，PEF 的阶数在

$$\frac{N}{3}-1<M<\frac{N}{2}-1 \tag{12-5-11}$$

间进行选择比较合适。

§12-6 自回归滑动平均(ARMA)谱估计

近几年来，利用自回归滑动平均模型(ARMA)进行功率谱估计的方法是继最大熵谱估计方法之后发展起来的方法。ARMA模型法是自回归(AR)法与滑动平均(MA)法的综合应用。这就是所谓极点零点模型法。用这种方法估计的谱的精度与分辨力均较好。

一、ARMA 模型

我们知道，一个线性随机过程可用下列线性滤波模型来描述

$$x_n=\mu+e_n+\psi_1 e_{n-1}+\psi_2 e_{n-2}+\cdots \tag{12-6-1}$$

其中e_n是均值为零方差为σ_e^2的白噪声序列，μ为描述过程x_n大小的量。假定线性滤波器是稳定的，这是确保过程x_n平稳的条件。此时，μ即为x_n的均值。进而设μ已被估出并已从时间序列中去掉，过程x_n就描述一个零均值的过程。

若在(12-6-1)式中取Z变换，$x(Z)=\Psi(Z)E(Z)$，并令$\Phi=\Psi^{-1}(Z)$，则有

$$\Phi(Z)X(Z)=E(Z) \tag{12-6-2}$$

在时间域上可以写成

$$x_n=\varphi_1 x_{n-1}+\varphi_2 x_{n-2}+\cdots+e_n \tag{12-6-3}$$

这里已经假定线性过程是可逆的，其条件是$\varphi(Z)$在单位圆及其内部收敛。若用(12-6-3)式中的有限项表示过程x_n，即

$$x_n=a_1 x_{n-1}+a_2 x_{n-2}+\cdots+a_p x_{n-p}+e_n \tag{12-6-4}$$

(12-6-4)式即为p阶AR过程的离散表示。这一表示式的涵义是：信号序列的现在值x_n可用一步预测法从它的过去值的有限的组合来获得，而e_n是误差。

对有限长度的滤波器，(12-6-1)式可以写成

$$x_n=e_n-\beta_1 e_{n-1}-\beta_2 e_{n-2}-\cdots-\beta_q e_{n-q} \tag{12-6-5}$$

这就称为q阶的滑动平均(MA)模型。

(12-6-4)和(12-6-5)式结合起来可以写成

$$\begin{aligned}x_n&=a_1 x_{n-1}+a_2 x_{n-2}+\cdots+a_p x_{n-p}+e_n\\&\quad-\beta_1 e_{n-1}-\beta_2 e_{n-2}-\cdots-\beta_q e_{n-q}\\&=\sum_{k=1}^{p}a_k x_{n-k}-\sum_{k=1}^{q}\beta_k e_{n-k}+e_n\qquad n=p+1,\cdots,N\end{aligned} \tag{12-6-6}$$

(12-6-6)式即为ARMA模型，或写为

$$x_n=-\sum_{k=1}^{p}a_k x_{n-k}+\sum_{k=0}^{q}b_k e_{n-k}\qquad n=p+1,\cdots,N \tag{12-6-7}$$

上式说明，一般时间序列可用确定的和非确定的两个部分来表示。前者用AR模型描述，后者用MA模型描述。其中e_n为$E[e_n]=0$和方差$E[e_n e_m]=\sigma_e^2\delta_{n,m}$的白噪声序列。

将(12-6-7)式两端取Z变换，可得到

$$A(Z)X(Z)=B(Z)E(Z) \tag{12-6-8}$$

式中

$$A(Z)=\sum_{k=0}^{p}a_k Z^{-k} \tag{12-6-8a}$$

$$B(Z)=\sum_{k=0}^{q}b_k Z^{-k} \tag{12-6-8b}$$

且有$a_0=1$。将$e(t)$视为系统的激励信号，ARMA模型的系统函数为

$$H(Z)=\frac{X(Z)}{E(Z)}=\frac{B(Z)}{A(Z)} \tag{12-6-9}$$

用$Z=e^{j2\pi fT}$代替Z，立即得到x_n的功率谱密度为

$$S_x(f)=\frac{B(e^{j2\pi fT})B(e^{-j2\pi fT})}{A(e^{j2\pi fT})B(e^{-j2\pi fT})}\sigma_e^2 \tag{12-6-10}$$

其中T为采样周期。

二、Pisarenko谐波分解法

若一过程只由一些正弦波和加性噪声组成，则可将它模型化为特殊情况下的ARMA过程。这种特殊情况下的数学特性导致一个求参数的特征值分析。因此它的处理又有别于一般的ARMA过程。

为了看出ARMA过程是白噪声加正弦波的合适模型，首先考虑一个三角恒等式

$$\sin(\Omega n) = 2\cos\Omega\sin[\Omega(n-1)] - \sin[\Omega(n-2)] \tag{12-6-11}$$

$-\pi \leq \Omega \leq \pi$，$\Omega = 2\pi fT$

记$x_n = \sin(\Omega n)$，则(12-6-11)式可表为

$$x_n = (2\cos\Omega)x_{n-1} - x_{n-2} \tag{12-6-12}$$

即现在的正弦值可由过去的两个值x_{n-1}、x_{n-2}来获得。将(12-6-12)式取Z变换后可得到

$$X(Z)[1 - 2\cos\Omega \cdot Z^{-1} + Z^{-2}] = 0 \tag{12-6-13}$$

多项式$1 - 2\cos\Omega Z^{-1} + Z^{-2}$在单位圆上的根（互为共轭）$Z_1 = \exp(j2\pi fT)$和$Z_2 = Z_1^* = \exp(-j2\pi fT)$就决定了正弦波的频率。而且有$f_1 = f_2$。

一般说，一个$2p$阶的形如

$$x_n = -\sum_{k=1}^{2p} a_k x_{n-k} \tag{12-6-14}$$

的实系数差分方程亦可表为由p个$\sin(2\pi f_i T)$形的实正弦所组成的确定过程。系数a_k为下列多项式的系数

$$Z^{2p} + a_1 Z^{2p-1} + \cdots + a_{p-1}Z^{p+1} + a_p Z^p + a_{p+1}Z^{p-1} + \cdots + a_{2p-1}Z + a_{2p}$$

$$= \sum_{i=1}^{p}(Z - Z_i)(Z - Z_i^*) \tag{12-6-15}$$

该多项式在单位圆上的根以复共轭对的形式出现。

对于正弦波加白噪声的情况，观察的序列为

$$y_n = x_n + v_n = -\sum_{i=1}^{2p} a_k x_{n-k} + v_n \tag{12-6-16}$$

其中有$E[v_n v_{n+k}] = \sigma_v^2 \delta_k$，$E[v_n] = 0$，又因噪声和正弦量不相关，故有$E[x_n v_n] = 0$。将$x_{n-k} = y_{n-k} - v_{n-k}$代入(12-6-16)式得到

$$\sum_{k=0}^{2p} a_k y_{n-k} = \sum_{k=0}^{2p} a_k v_{n-k} \tag{12-6-17}$$

可见上式是属于ARMR模型，只是AR和MA的阶数皆为$2p$。

(12-6-17)式的矩阵形式为

$$\boldsymbol{y}^T \boldsymbol{A} = \boldsymbol{v}^T \boldsymbol{A} \tag{12-6-18}$$

其中

$$\boldsymbol{y}^T = [y_n, y_{n-1}, \cdots, y_{n-2p}]$$

$$\boldsymbol{A}^T = [1, a_1, a_2, \cdots, a_{2p-1}, a_{2p}]$$

$$\boldsymbol{v}^T = [v_n, v_{n-1}, \cdots, v_{n-2p}]$$

(12-6-18)式两边乘以$\boldsymbol{y}$并取期望值，得到

$$E[\boldsymbol{y}\boldsymbol{y}^T]\boldsymbol{A} = E[\boldsymbol{y}\boldsymbol{v}^T]\boldsymbol{A}$$

$$= E[(\boldsymbol{x} + \boldsymbol{v})\boldsymbol{v}^T]\boldsymbol{A}$$

$$= E[\boldsymbol{V}\boldsymbol{v}^T]\boldsymbol{A} = \sigma_v^2 \boldsymbol{A} \tag{12-6-19}$$

式中已定义$\boldsymbol{x} = [x_n, \cdots, x_{n-2p}]^T$。进一步记序列$y_n$的协方差矩阵$E[\boldsymbol{y}\boldsymbol{y}^T] = \boldsymbol{R}_{yy}$，则(12-6-19)式可写为

$$R_{yy}A = \sigma_v^2 A \qquad (12\text{-}6\text{-}20)$$

(12-6-20)式是一个特征方程，而噪声方差 σ_v^2 是 R_{yy} 的一个特征值。ARMA 参数矢量 A 是关于特征值 σ_v^2 的特征矢量。当已知 $2p+1$ 个自相关值时，可解出 A，ARMA 模型中的谐波频率由(12-6-15)式所示的多项式的根来确定。

(12-6-20) 式的特征矢量不包含各谐波的幅度信息，须利用观察序列 y_n 的协方差来求。对于白噪声中的正弦波有

$$r_{yy}(0) = \sigma_v^2 + \sum_{i=1}^{p} P_i$$

$$r_{yy}(k) = \sum_{i=1}^{p} P_i \cos(2\pi f_i kT), k \neq 0 \qquad (12\text{-}6\text{-}21)$$

(12-6-21)式写为矩阵形式有

$$FP = r \qquad (12\text{-}6\text{-}22)$$

其中

$$F = \begin{bmatrix} \cos(2\pi f_1 T) & \cdots & \cos(2\pi f_p T) \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \cos(2\pi f_1 pT) & \cdots & \cos(2\pi f_p pT) \end{bmatrix}$$

$$P = [P_1, P_2, \cdots, P_p]^T$$

$$r = [r_{yy}(1), r_{yy}(2), \cdots, r_{yy}(p)]^T$$

由(12-6-22)式可解出各正弦波的功率。噪声功率可由下式确定

$$\sigma_v^2 = r_{yy}(0) - \sum_{i=1}^{p} P_i$$

由上分析可知，Pisarenko的谐波分解法是ARMA模型的特例。它的主要关键是在已知输入序列 y_n 的头 $2p+1$ 个自相关值的条件下，求解特征方程(12-6-20)，以确定模型的系数，从而由(12-6-15)式的多项式的根求得各谐波分量的频率。

三、ARMA模型参数的确定

重写(12-6-7)式如下

$$x_n = -\sum_{k=1}^{p} a_k x_{n-k} + \sum_{k=0}^{q} b_k e_{n-k} \qquad n = p+1, \cdots, N \qquad p \geqslant q$$

用 x_{n-l} 乘上式两端，然后取期望值得到

$$r_{xx}(l) = -\sum_{k=1}^{p} a_k r_{xx}(l-k) + \sum_{k=0}^{q} b_k r_{ex}(l-k) \qquad (12\text{-}6\text{-}23)$$

其中

$$r_{ex}(k) = E[e_n x^*_{n-k}]$$

由于将来的输入不会影响一个因果稳定滤波器的输出，故有 $r_{ex}(k) = 0$（当 $k>0$）。因此有

$$r_{xx}(l) = \begin{cases} -\sum_{k=1}^{p} a_k r_{xx}(l-k) + \sum_{k=0}^{q} b_k r_{ex}(l-k), & l = 0, \cdots, q \\ -\sum_{k=1}^{p} a_k r_{xx}(l-k), & l = q+1, q+2, \cdots \end{cases}$$

又注意到

$$r_{ex}(k)=E[e_{n+k}x_n^*]$$

$$=E\left[e_{n+k}\sum_{i=0}^{\infty}h_i^*e_{n-i}^*\right]$$

$$=\sum_{i=0}^{\infty}h_i^*\sigma^2\delta_{k+i}=\sigma_e^2h_{-k}^*$$

其中h_k为系统冲激响应，因此(12-6-23)式可写成

$$r_{xx}(l)=\begin{cases}-\sum_{k=1}^{p}a_kr_{xx}(l-k)+\sigma^2\sum_{k=l}^{q}b_kh^*_{k-l}, & l=0,1,\cdots,q \quad (12\text{-}6\text{-}24)\\ -\sum_{k=1}^{p}a_kr_{xx}(l-k), & l=q+1,q+2,\cdots \quad (12\text{-}6\text{-}24a)\end{cases}$$

方程(12-6-24a)常称为修正的Yule-Walker方程。因此，确定 ARMA 模型参数的一般方法分两步。首先利用(12-6-24a)式的修正Yule-Walker方程求AR模型参数a_k。这一步是与MA参数无关的。然后利用(12-6-24)式确定AM 模型参数 b_k 。一般说第二步常需要解非线性方程组。这是因为冲激响应是a_k、b_k 的函数。

对于只是进行功率谱估计的场合，一般不必去求MA模型参数。由(12-6-10)式可知，此时只要求出$A(Z)A(1/Z)$和 $B(Z)B(1/Z)$ 便可。不失一般性，假设 $\sigma_e^2=1$，令B_k 为$B(Z)B(1/Z)$的逆Z 变换所得的序列，记为$B_k\Longleftrightarrow B(Z)B(1/Z)$，则(12-6-10)式的谱估值为

$$S_x(f)=\sum_{k=-q}^{q}B_k\exp(-j2\pi fkT)/|A\exp(j2\pi fT)|^2 \qquad (12\text{-}6\text{-}25)$$

其中$B_{-k}=B_k^*$为求得B_k，记$A_k\Longleftrightarrow A(Z)A(1/Z)$，另一方面由(12-6-10)式可知 $B_k\Longleftrightarrow A(Z)A(1/Z)\cdot S_x(Z)$加上对于自相关序列有$r_{xx}(k)\Longleftrightarrow S_x(Z)$。由$Z$变换的特性容易得知，$B_k$是$A_k$和$r_{xx}(k)$的卷积，即

$$B_k=\sum_{l=-p}^{p}A_lr_{xx}(k-l) \qquad k=0,1,\cdots,q \qquad (12\text{-}6\text{-}26)$$

另一种估计 ARMA 模型参数的方法是先用一个AR模型来估计x_n，然后再设法找出这一AR模型的参数与ARMA模型参数的关系，从而求出ARMA模型的参数。

我们首先用一个无限阶的AR模型去和ARMA模型等同，亦即令

$$C(Z)=1+\sum_{k=1}^{\infty}c_kZ^{-k}=\frac{E(Z)}{X(Z)} \qquad (12\text{-}6\text{-}27)$$

其中c_k即为输人序列x_n的AR模型参数。注意到(12-6-9)式，因而亦有

$$\hat{C}(Z)=\frac{A(Z)}{B(Z)} \qquad (12\text{-}6\text{-}28)$$

我们用一个$M\geqslant p+q$阶的AR模型

$$\hat{C}(Z)=1+\sum_{k=1}^{M}\hat{c}^kZ^{-k} \qquad (12\text{-}6\text{-}29)$$

的参数来作为估计的AR参数。设$p>q$，有

$$\widehat{C}(Z)=\frac{\widehat{A}(Z)}{\widehat{B}(Z)} \tag{12-6-30}$$

或

$$\widehat{B}(Z)\widehat{C}(Z)=\widehat{A}(Z) \tag{12-6-30a}$$

由Z变换特性知$\widehat{B}(Z)\widehat{C}(Z)$的逆Z变换是序列$\widehat{b}_n$和$\widehat{c}_n$的卷积，即

$$\sum_{k=0}^{q}\widehat{b}_k\widehat{c}_{n-k}=\widehat{a}_n \qquad n=1,2,\cdots \qquad \widehat{b}_0=0 \tag{12-6-31}$$

由于当$n>p$时，a_n应当为零，故(12-6-31)式写为

$$\sum_{k=0}^{q}\widehat{b}_k\widehat{c}_{n-k}=0 \qquad n=p+1,p+2,\cdots,p+q \tag{12-6-31a}$$

$$\sum_{k=0}^{q}\widehat{b}_k\widehat{c}_{n-k}=\widehat{a}_n \qquad n=1,2,\cdots,p \tag{12-6-31b}$$

(12-6-31a)和(12-6-31b)均为线性方程组。由(12-6-31a)可解得$\widehat{b}_n(n=1,2,\cdots,q)$，然后由(12-6-31b)式立即解出$\widehat{a}_n(n=1,2,\cdots,p)$。至此，$\widehat{a}_n$和$\widehat{b}_n$都已得到，从而利用(12-6-10)式可得到$x_n$的谱估值。由于$C(Z)=A(Z)/B(Z)$，当$B(Z)$的零点靠近单位圆时，必须采用大阶数的AR模型。

Konvalenka等人提出把ARMA模型化为最小二乘输入输出辨识问题。这里介绍一下这一方法的要点。由(12-6-23)式可知，正规方程的非线性是由未知的输入和输出互相关函数$r_{ex}(k)$所造成的。如e_n不可观测，$R_{ex}(k)$就无法估计。但若e_n已知，则ARMA参数就可由一组线性方程解得。

为建立线性方程组，将(12-6-7)式改写如下

$$x_n=-\sum_{k=1}^{p}a_kx_{n-k}+\sum_{k=1}^{q}b_ke_{n-k}+e_k \qquad n=0,1,\cdots,N-1$$

写成矩阵形式有

$$\boldsymbol{x}=\Phi\theta+\boldsymbol{e} \tag{12-6-32}$$

其中

$$\boldsymbol{x}=[x_0,x_1,\cdots,x_{N-1}]^T$$
$$\boldsymbol{e}=[e_0,e_1,\cdots,e_{N-1}]^T$$
$$\theta=[-a_1,-a_2,\cdots-a_p,b_1,b_2,\cdots,b_q]^T$$

以及

$$\Phi=\begin{pmatrix} x_{-1} & x_{-2} & \cdots & x_{-p} & e_{-1} & e_{-2} & \cdots & e_{-q} \\ x_0 & x_{-1} & \cdots & x_{-p+1} & e_0 & e_{-1} & \cdots & e_{-q+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{N-1} & x_{N-2} & \cdots & x_{N-p-1} & e_{N-2} & e_{N-3} & \cdots & e_{N-q-1} \end{pmatrix}$$

这里Φ为$N\times(p+q)$维输入-输出数据矩阵。(12-6-32)式是标准的线性最小二乘估计问题其解为

$$\widehat{\theta}=(\Phi^H\Phi)^{-1}\Phi^H\boldsymbol{x} \tag{12-6-33}$$

它与AR过程的最小二乘参数估计方法相似，只是相关矩阵$R=\Phi^H\Phi$包括互相关值$r_{ex}(k)$。初始条件$\{x_{-p},x_{-p+1},\cdots,x_{-1},e_{-q},e_{-q+1},\cdots,e_{-1}\}$可假定为零。为了估计$r_{ex}(k)$，先要估计$e_n$。可用大阶数的AR模型来估计$x_n$，$e_n$就等于预测误差时间序列。根据$e_n$可得到$r_{ex}(k)$的估值，从而利用(12-6-33)式求得ARMA模型参数$\hat{\theta}$。

近年来，如何估计ARMA模型参数仍是很受研究者注意的问题。大多数已发展的方法，在性能和计算复杂性方面似乎没有什么明显的差别。

§12-7 最小交叉熵谱估计方法

一、引言

最大熵和最小交叉熵谱估计方法都是信息论的推理方法。一般的推理问题可这样提出：假定我们知道一个系统有一组可能的状态x_i，它们的概率密度函数$q^+(x_i)$未知。我们又知道关于分布q^+的一些约束——某些期望值$\sum_i q^+(x_i)f_k(x_i)$。现在要选择一个分布q，使它在某种意义上是q^+的最好估计，我们应当选择哪一个q？

最大熵原理表明，在所有满足约束条件的分布q中应当选择使熵$-\sum_i q(x_i)\log(q(x_i))$最大的一个。而最小交叉熵原理则说明，除了上述约束之外，当已知估计q^+的一个先验密度p时，应当选择使交叉熵$\sum_i q(x_i)\log(q(x_i)/p(x_i))$最小的那一个$q$。下面将会看到，这两种方法有密切联系。如果先验密度p是均匀分布时，最小交叉熵就是最大熵。对于连续变量的情况，交叉熵定义为

$$H(q,p)=\int q(x)\log(q(x)/p(x))dx \tag{12-7-1}$$

于是问题的一般叙述如下：

设x为某系统的一个状态，可能状态的集合为D，这个状态的概率密度函数为$q^+(x)$假定D为使得$q(x)\geqslant 0$（对$x\in D$）的全部概率密度q在D内的集合，且有

$$\int_D q(x)dx=1 \tag{12-7-2}$$

我们只知道“真”密度$q^+(x)\in$ D，但q^+本身是未知的。

设 $p\in$ D是一个先验密度，这是 q^+ 的一个现行估值。并假定，对一个已知的函数集$g_r(x)$和数集$\bar{g}_r(r=0,1,\cdots M)$，我们以下列期望值形式得到了关于$q^+$的新信息

$$\int_D q^+(x)g_r(x)dx=\bar{g}_r,\ \ r=0,1,\cdots,M \tag{12-7-3}$$

由于约束条件（12-7-3）并不能完全确定q^+，故而不仅q^+，而且某些密度函数子集L $\in$ D亦满足这些条件。问题变为应当怎样选择q^+的估值以及如何利用先验密度p作出这一选择。文献[16]证明了以(12-7-3)式的期望值形式给出新信息的约束条件下，使（12-7-1）式的交叉熵对先验密度$p(x)$求最小值来选择的后验密度$q(x)$是问题的唯一解。亦即后验密度满足条件

$$H(q,p)=\min H(q',p)\qquad q'\in \mathrm{L} \tag{12-7-4}$$

其中L $\subset$ D包含满足约束条件(12-7-3)的全部概率密度。

应用拉格朗日乘子法，容易求得满足约束条件(12-7-3)和(12-7-2)时，使交叉熵最小的条件为

$$\log \frac{q(x)}{p(x)} + 1 + \lambda_0 + \sum_{r=0}^{M} \beta_r g_r(x) = 0 \qquad (12\text{-}7\text{-}5)$$

其中β_r和λ_0分别为对应于约束条件(12-7-3)和(12-7-2)的拉格朗日乘子。

由(12-7-5)式可求得$q(x)$的解为

$$q(x) = p(x)\exp\left(-\lambda - \sum_{r=0}^{M} \beta_r g_r(x)\right) \qquad (12\text{-}7\text{-}6)$$

式中$\lambda = \lambda_0 + 1$。，若能求得满足约束条件(12-7-3)和(12-7-2)的β_r和λ，则上式即为所求得的$q(x)$。

若记$z = \exp(\lambda)$，将(12-7-6)式代入约束条件(12-7-2)式，可得

$$z = \int p(x)\exp\left(-\sum_{r=0}^{M} \beta_r g_r(x)\right) dx \qquad (12\text{-}7\text{-}7)$$

再将(12-7-6)式代入(12-7-3)式，得到

$$\bar{g}_r = z^{-1}\int p(x) g_r(x)\exp\left(-\sum_{k=0}^{M} \beta_k g_k(x)\right) dx \qquad (12\text{-}7\text{-}8)$$

其中z由(12-7-7)式确定，于是原则上可由(12-7-8)式解得β_r，实际上直接利用期望值$\bar{g}_r$得到β_r的解析解仍是相当困难的。

二、最小交叉熵谱

我们将用上述最小交叉熵原理来估计随机信号的功率谱密度。假定时间信号是一个平稳随机过程 $g(t)$，因为任一平稳随机过程皆可以表为一系列离散频率正余弦过程之和的极限，故首先考虑包括有限项和的表示式

$$s(t) = \sum_{k=1}^{N}\left[a_k\cos(2\pi f_k t) + b_k\sin(2\pi f_k t)\right] \qquad (12\text{-}7\text{-}9)$$

其中a_k与b_k均为随机变量，f_k是不为零的频率。适当选择a_k和b_k以及频率f_k，可以使均方误差$E[|g(t) - s(t)|^2]$最小。因每个频率点的功率可用变量 $x_k = \frac{1}{2}(a_k^2 + b_k^2)$表示，所以可用$x_1, x_2, \cdots, x_N$的联合概率密度函数$q(x)$来描述上述随机过程。

令$S_k, k = 1, \cdots, N$为包含未知过程g^+功率谱的一个先验估计。此时，假定$q^+(x)$的先验估计$p(x)$具有下列指数形式

$$p(x) = \prod_{k=1}^{N}(1/S_k)\exp[-x_k/S_k] \qquad (12\text{-}7\text{-}10)$$

因为

$$\int x_k p(x)\,dx = S_k$$

故在频率f_k的期望功率正好就是先验估计S_k。利用变量a_k和b_k表示，上式变成

$$p(a,b) = \prod_{k=1}^{N}\left(\frac{1}{2\pi S_k}\right)\exp[-(a_k^2 + b_k^2)/2S_k] \qquad (12\text{-}7\text{-}11)$$

这表明(12-7-10)式的先验估计假设等价于假定 a_k、b_k 为高斯分布变量。这种假定通常是合理的，且数学处理亦方便。

现在我们令 T_k 为 $q^+(x)$ 的真正功率谱密度，即

$$T_k = \int x_k q^+(x)dx \tag{12-7-12}$$

并假定由自相关函数 $r(t_r)$ 的 $M+1$ 个值得到关于 q^+ 的新信息，即

$$R_r = r(t_r) = \sum_{k=-N}^{N} T_k \exp(2\pi j f_k t_r) \qquad r = 0,1,\cdots,M \tag{12-7-13}$$

且有 $t_0 = 0$。上式用实数形式表示，可写为

$$R_r = \sum_{k=1}^{N} T_k c_{rk} \tag{12-7-14}$$

其中 $c_{rk} = 2\cos(2\pi t_r f_k)$。将(12-7-12)式代入(12-7-14)式，得到

$$R_r = \int\left(\sum_{k=1}^{N} x_k C_{rk}\right) q^+(x)dx \tag{12-7-15}$$

上式具有期望约束条件(12-7-3)式的形式。当给出了(12-7-10)式的先验估值及(12-7-15)式的新息之后，可由(12-7-6)式得到 q^+ 的最小交叉熵后验估值为

$$q(x) = p(x)\exp\left(-\lambda - \sum_{r=0}^{M}\beta_r \sum_{k=1}^{N} x_k c_{rk}\right) \tag{12-7-16}$$

其中 β_r 和 λ 分别为由(12-7-15)式及(12-7-2)式的约束条件所决定的拉格朗日乘子。将(12-7-10)式代入(12-7-16)式，得到

$$q(x) = \exp(-\lambda)\prod_{k=1}^{N}\left\{(1/S_k)\exp\left[-\left(\frac{1}{S_k} + \sum_{r=0}^{M}\beta_r c_{rk}\right)x_k\right]\right\}$$

$$= \exp(-\lambda)\prod_{k=1}^{N}\left\{(1/S_k)\exp\left[-\left(\frac{1}{S_k} + U_k\right)x_k\right]\right\} \tag{12-7-17}$$

其中

$$U_k = \sum_{r=0}^{M}\beta_r c_{rk} \tag{12-7-18}$$

利用约束条件(12-7-2)式，可得

$$\exp(\lambda) = \int\prod_{k=1}^{N}\left\{(1/S_k)\exp\left[-\left(\frac{1}{S_k} + U_k\right)x_k\right]\right\}dx$$

$$= \prod_{k=1}^{N}\left[(1/S_k)\frac{1}{\frac{1}{S_k} + U_k}\right] \tag{12-7-19}$$

将(12-7-19)式代入(12-7-17)式立即得到 x 的密度函数的后验估计为

$$q(x) = \prod_{k=1}^{N}\left(\frac{1}{S_k} + U_k\right)\exp\left[-\left(\frac{1}{S_k} + U_k\right)x_k\right] \tag{12-7-20}$$

因此，后验估计也是形如(12-7-10)式的指数分布。功率谱的后验估值正好是

$$T_k = \int x_k q(x) dx$$
$$= \frac{1}{\frac{1}{S_k} + U_k} = \frac{S_k}{1 + S_k \sum_{r=0}^{N} \beta_r c_{rk}} \qquad (12\text{-}7\text{-}21)$$

其中乘子β_r应选择得使T_k满足自相关约束(12-7-14)式，

(12-7-21) 式可用线性滤波器来解释。一个先验密度为$p(x)$的过程通过特性函数为$Y(f)$的线性滤波器，每一功率谱分量的幅度变化了一个因子 $A_k = |Y(f_k)|^2$，从S_k变成$S_k A_k$。因此(12-7-21)式表明，后验密度$q(x)$可以利用先验密度函数通过$A_k = (1 + S_k v_k)^{-1}$的线性滤波器来获得。

利用这一观点可知，滤波器输入$p(x)$与输出$q(x)$之间的关系为

$$q(x) = \frac{p(x_1/A_1, x_2/A_2, \cdots, x_N/A_N)}{A_1 A_2 \cdots A_N} \qquad (12\text{-}7\text{-}22)$$

由(12-7-1)式可求得$p(x)$与$q(x)$的交叉熵为

$$H(p,q) = \int q(x) \log(q(x)/p(x)) dx$$
$$= \int p(y_1, y_2, \cdots, y_N) \log\left[\frac{p(y_1, y_2, \cdots, y_N)}{p(y_1 A_1, y_2 A_2, \cdots, y_N A_N)}\right] dy$$
$$- \sum_{k=1}^{N} \log(A_k) \qquad (12\text{-}7\text{-}23)$$

式中利用了$y_k = x_k/A_k$及$dy = dx / \prod_{k=1}^{N} A_k$两个关系式。

(12-7-23)式是输入$p(x)$为任意形式时求交叉熵的一般公式。

三、最小交叉熵谱估计与最大熵谱估计间的关系

当没有谱形状的先验信息可资利用时，合适的选择是假定输入密度函数 $p(x)$ 是均匀的。此时，(12-7-23)式中第一项为零，因而滤波器输出与输入间的交叉熵为

$$H(p,q) = -\sum_k \log(A_k) \qquad (12\text{-}7\text{-}24)$$

这是输入过程熵的离散形式，只是多一个负号（比较(12-1-7b)式）。因此，使(12-7-24)式最小就等价于使过程的熵最大，从而证明了最大熵谱是当先验分布为均匀分布时最小交叉熵谱的特例。

若先验谱形状信息为(12-7-10)式的指数分布（谱幅度 a_k、b_k 为高斯分布量），而且先验谱估值是所谓"白谱"，即$S_k = S$，$k = 1, 2, \cdots, N$。应用(12-7-21)式有

$$T_k = \frac{1}{\sum_{r=0}^{N} \beta_r' c_{rk}} \qquad (12\text{-}7\text{-}25)$$

式中β_r'是修正的拉格朗日乘子，满足

$$\beta_0' = \beta_0 + \frac{1}{S}$$

$$\beta_r' = \beta_r \qquad r = 1, 2, \cdots, M$$

若自相关函数延迟是等间隔的，即$t_r = r\Delta t$（$t_0 = 0$），则可将(12-7-25)式写为

$$T(f) = \frac{1}{\sum_{r=-M}^{M} \beta_r' Z^{-r}} \tag{12-7-26}$$

其中$Z = \exp(j2\pi f \Delta t)$，并已省去谱估值下标$k$。

运用Fejèr-Riesz定理[20]，(12-7-26)式可写为(或仿照(12-1-19)式的推导过程)

$$\hat{T}(f) = \frac{\sigma^2}{\left|\sum_{r=0}^{M} a_r Z^{-r}\right|^2} \tag{12-7-27}$$

其中$a_0 = 1$，σ^2 为增益，a_r是逆滤波器的系数。这一表示式与最大熵谱有相同的形式。比较(12-7-27)式和(12-7-26)式，可得到β_r'与a_r的关系如下

$$\beta_r' = \frac{1}{\sigma^2} \sum_{k=1}^{M-r} a_{k+r} a_k^* \tag{12-7-28}$$

由上分析可知，最大熵与最小交叉熵方法的基本区别在于后者利用了先验信息。因而利用最小小交叉熵准则得到的谱估值比用最大熵准则得到的谱优越，亦即谱的分辨力较高。但是最小交叉熵谱估计适用的信号是振幅正态分布的互相独立的多正弦信号，因而这种模型不如AR模型适用范围广。其次，它不能象最大熵谱那样直接依据记录得到谱估计，而必须从数据记录得到自相关估计，进而获得最小交叉熵谱，从而给使用带来不便。更重要的是待定的乘子难于求得。由(12-7-14)式和(12-7-21)式可知，为求待定乘子β_r，需要解下列方程

$$\sum_{k=1}^{N} \left(\sum_{r=0}^{M} \beta_r c_{rk} + S_k^{-1} \right)^{-1} c_{rk} = R_r \tag{12-7-29}$$

这是一个非线性方程组。文献[18]给出了非线性迭代法，但计算是相当复杂的。因此，最小交叉熵谱估计技术的广泛应用受到一定的限制，国内外学者都在研究，以期得到各种有效算法。

§12-8 最大似然谱估计

最大似然谱估计方法首先由Capon用于地震线列阵的波数谱估计。这一方法也可用于单一信号的谱估计。应当指出，最大似然谱估计只是一种误称，因为人们只是通过有效地测量窄带滤波器输出功率来估计功率谱密度。这样获得的谱估值并非真正的功率谱的最大似然估计，只是一种最小方差无偏估计或最小能量估计。由估计理论我们知道，信号的最大似然估计在高斯分布的条件下也是最小方差无偏估计。因此本节在讨论最大似然谱估计时只给出最小方差无偏估计的推导。

但是，我们知道，最大似然法的基本原理是将有限时间内的随机序列独立观测数据样本$x_n (n = 1, 2, \cdots, N)$进行分析，将$N$个样本值的联合概率密度函数视为待估参数的函数$\theta$，然后使获得那些观测值的概率为极大来求得所需的参数估值。因而AR模型参数亦可用最

大似然法来进行估计，从而得到 AR 谱估计。这种方法称为最大似然AR参数估计。作为模型参数的估计与最大似然谱估计尽管放在同一节加以讨论，但由上面的讨论可知这两者是不同的。

一、最大似然谱估计

这里我们给出最小方差无偏估计的推导，并解释为什么这一估计是一个功率谱密度的合适估计。

假定一个离散时间序列$y(n)$通过一个长为M的线性滤波器之后的输出序列为

$$x(k)=\sum_{m=1}^{M}a_m y(k-m) \qquad k=M+1,\cdots,N \tag{12-8-1}$$

其中a_m为滤波器的系数。若a_m及输入皆为复数，即

$$y(k)=Ae^{j2\pi fk}+v(k) \tag{12-8-2}$$

而$v(k)$为与频率f及幅度A无关的零均值随机过程。我们可以这样选择一组a_m，使滤波器通过分量$A\exp[j2\pi fk]$，而排斥$v(k)$。

若$x(k)$是$A\exp[j2\pi fk]$的无偏估计，则有

$$A\exp[j2\pi fk]=\sum_{m=1}^{M}a_m A\exp[j2\pi f(k-m)] \tag{12-8-3}$$

或者

$$1=\sum_{m=1}^{M}a_m\exp[j2\pi f(-m)] \tag{12-8-3a}$$

我们采用记号

$$E=[e^{j2\pi f},e^{j2\pi 2f},\cdots,e^{j2\pi Mf}]^T \tag{12-8-4}$$

则(12-8-3)式的约束条件可写为

$$E^H a=I \tag{12-8-5}$$

或

$$a^H E=I \tag{12-8-5a}$$

式中$a=[a_1,a_2,\cdots,a_M]^T$记号H表示共轭转置，I为单位矩阵。估计量的方差为

$$\begin{aligned}\sigma^2&=E[x(k)-E(x(k))]^2\\&=E\left[\sum_{m=1}^{M}a_m v(k-m)\right]^2\\&=\sum_{n=1}^{M}\sum_{m=1}^{M}a_n^* a_m r(n,m)\\&=a^H R a\end{aligned} \tag{12-8-6}$$

式中R为噪声过程$v(n)$的$M\times M$维相关矩阵。现在要求出在约束条件(12-8-5)式上使方差σ^2最小的条件，由(12-8-5)和(12-8-5a)式可知有

$$a^H E E^H a=1 \tag{12-8-7}$$

或者

$$a^H \Phi a=1 \tag{12-8-7a}$$

成立，其中$\Phi=EE^H$。于是可作函数

$$F = a^H R a + \lambda(1 - a^H \Phi q) \qquad (12\text{-}8\text{-}8)$$

式中λ为拉格朗日乘子。σ^2极小值应满足的条件为

$$\frac{\partial F}{\partial a} = 2Ra - \lambda 2\Phi a = 0 \qquad (12\text{-}8\text{-}9)$$

由此不难求得

$$a = \lambda R^{-1} E \qquad (12\text{-}8\text{-}10)$$

将(12-8-10)式代入约束条件(12-8-5)式，可解得乘子如下

$$\lambda = \frac{1}{E^H R^{-1} E} \qquad (12\text{-}8\text{-}11)$$

由(12-8-10)及(12-8-11)式可得到系数a为

$$a = \frac{R^{-1} E}{E^H R^{-1} E} \qquad (12\text{-}8\text{-}12)$$

将(12-8-12)式代入(12-8-6)式，从而得到滤波器输出功率的最小功率为

$$\sigma^2 = \frac{1}{E^H R^{-1} E} \qquad (12\text{-}8\text{-}13)$$

若滤波器输入过程$y(t)$是零均值随机过程，则 R 即为该过程的相关矩阵。此时，若这样设计滤波器，使其在所考虑的频率f上的频率响应为1（即满足约束条件(12-8-3a)式），则输出最小方差由(12-8-13)式决定。在这种情况下，滤波器无失真地通过频率 f 及其附近的频率分量，而抑制了不在f 附近的分量。亦即输出功率主要是由接近f 的频率分量所贡献的。因此(12-8-13) 式反映了输入过程的功率谱估计。正是因为这个原因，有时把这种谱估计方法称为最小能量法。

二、最大似然AR参数估计

如本章§12-6所述，一个随机时间序列$x(n)$的AR模型可表为

$$x(k) = \sum_{m=1}^{M} a_m x(k-m) + v(k) \qquad k = M+1, \cdots, N \qquad (12\text{-}8\text{-}14)$$

其中$v(k)$是误差。输入过程的AR谱估计为

$$S(f) = \frac{\sigma^2_M}{\left|1 + \sum_{m=1}^{M} a_m e^{-j2\pi f m T}\right|^2} \qquad (12\text{-}8\text{-}15)$$

现在我们用最大似然法来估计参数a_m和误差的方差 σ^2_M。

考虑似然函数

$$\begin{aligned} p(x_1, \cdots, x_N / a_1, \cdots, a_M, \sigma^2_M) &= p(x_{M+1}, \cdots, x_N / x_1, \cdots, x_M, a_1, \cdots, a_M, \sigma_M^2) \\ &\quad \times p(x_1, \cdots, x_M / a_1, \cdots, a_M, \sigma^2_M) \end{aligned} \qquad (12\text{-}8\text{-}16)$$

通常序列长度N总是比AR模型阶数M大得多，故少数(M) 个变量是否已知，并不显著改变大量随机变量的概率分布。因而$x_1, \cdots, x_N$的联合概率分布可用(12-8-16)式中第一因子近似，即

$$p(x_1,\cdots,x_N/a_1,\cdots,\sigma^2_M)$$
$$\approx p(x_{M+1},\cdots,x_N/x_1,\cdots,x_M,a_1,\cdots,a_M,\sigma_M^2) \qquad (12\text{-}8\text{-}17)$$

设$v(k)$为白高斯序列，则$v_{M+1},\cdots,v_N$的联合概率密度函数为

$$p(v_{M+1},\cdots,v_N)$$
$$\doteq \frac{1}{(2\pi\sigma^2_M)^{(N-M)/2}}\exp\left\{-\frac{1}{2\sigma^2_M}\sum_{n=M+1}^{N}v_n^2\right\} \qquad (12\text{-}8\text{-}18)$$

由(12-8-14)式可知，在给定$x_1,\cdots,x_M$条件下，序列$x_{M+1},\cdots,x_N$是由$v_{M+1},\cdots,v_N$作线性变换得到的。这一线性变换可记为

$$\boldsymbol{v}=\boldsymbol{X}\boldsymbol{a} \qquad (12\text{-}8\text{-}19)$$

其中

$$\boldsymbol{v}=[v_{M+1},\cdots,v_N]^T \qquad (12\text{-}8\text{-}20a)$$

$$\boldsymbol{X}=\begin{pmatrix} x_{M+1} & -x_M & \cdots & x_1 \\ x_{M+2} & -x_{M+1} & \cdots & -x_2 \\ \cdots & \cdots & \cdots & \cdots \\ x_N & -x_{N-1} & \cdots & -x_{N-M} \end{pmatrix} \qquad (12\text{-}8\text{-}20b)$$

$$\boldsymbol{a}=[1,a_1,\cdots,a_M]^T \qquad (12\text{-}8\text{-}20c)$$

容易得知，变换为雅可比

$$\left|\frac{\partial(v_{M+1},\cdots,v_N)}{\partial(x_{M+1},\cdots,x_N)}\right|=1 \qquad (12\text{-}8\text{-}21)$$

因此(12-8-16)式的联合概率密度函数可近似为

$$p(x_1,\cdots,x_N/a_1,\cdots,a_M,\sigma^2_M)$$
$$\approx(2\pi\sigma^2_M)^{-(N-M)/2}$$
$$\times\exp\left\{-\frac{1}{2\sigma_M^2}\sum_{j=M+1}^{N}(x_j-a_1x_{j-1}-\cdots-a_Mx_{j-M})^2\right\} \qquad (12\text{-}8\text{-}22)$$

我们把已观测到的样本值的联合概率密度函数看作是待估计量$a_1,\cdots,a_M\sigma_M^2$的函数。于是可将(12-8-18)式的似然函数取对数，再对$a_k(k=1,\cdots,M)$，σ_M^2求导，并令其等于零，亦即

$$\frac{\partial}{\partial a_k}\ln p(x_1,\cdots,x_N/a_1,\cdots,a_M,\sigma_M^2)=0 \qquad k=1,\cdots,M \qquad (12\text{-}8\text{-}23)$$

$$\frac{\partial}{\partial\sigma_M^2}\ln p(x_1,\cdots,x_N/a_1,\cdots,a_M,\sigma_M^2)=0 \qquad (12\text{-}8\text{-}24)$$

从而得到各参数的最大似然估值应满足的方程

$$\sum_{j=M+1}^{N}\sum_{i=1}^{M}\hat{a}_i x_{j-i}x_{j-k}=\sum_{j=M+1}^{N}x_jx_{j-k} \qquad (12\text{-}8\text{-}25)$$

和

$$\hat{\sigma}^2_M=\frac{1}{N-M}\sum_{j=M+1}^{N}(x_j-a_1x_{j-1}-\cdots-a_Mx_{j-M})^2 \qquad (12\text{-}8\text{-}26)$$

(12-8-25)式左端交换求和次序之后可写为

$$\sum_{l=1}^{M}\hat{a}_l\hat{r}_{xx}(l-k)=\hat{r}_{xx}(k)\qquad k=1,\cdots,M \tag{12-8-27}$$

其中

$$\hat{r}_{xx}(l-k)=\sum_{i=M+1}^{N}x_{i-l}x_{i-k} \tag{12-8-28}$$

$$\hat{r}_{xx}(k)=\hat{r}_{xx}(-k) \tag{12-8-29}$$

为相关函数$r_{xx}(k)$的估值。解线性方程组(12-8-27)式便可得出$a_1,\cdots,a_M$的估值，再利用(12-8-26)式求得σ^2_M的估值$\hat{\sigma}^2_M$，从而由(12-8-15)式得到序列$x(n)$的功率谱估值。比较(12-8-27)式与本章§12-2中(12-2-6)式，可以得知，利用最大似然法估计AR参数的方程与预测滤波方程的区别是前者用相关函数的估值$\hat{r}_{xx}(k)$代替了相关函数$r_{xx}(k)$。实际上，当$N-M$足够大时(12-8-27)式中的$\hat{r}_{xx}(k)$可代之以真实相关函数$r_{xx}(k)$。因而可以得知，估计AR参数的最大似然法当$N-M$很大时与一步预测误差滤波法是等效的。

附录A 施瓦兹不等式

设$g_1(t)$和$g_2(t)$是区间$a \leqslant t \leqslant b$中的两个实函数，并满足以下条件

$$\int_a^b |g_1(t)|^2 dt < \infty \tag{A-1}$$

$$\int_a^b |g_2(t)|^2 dt < \infty \tag{A-2}$$

则存在下列施瓦兹不等式

$$\left|\int_a^b g_1(t) g_2(t) dt\right|^2 \leqslant \left(\int_a^b |g_1(t)|^2 dt\right)\left(\int_a^b |g_2(t)|^2 dt\right) \tag{A-3}$$

现在来证明此不等件，构造一个新函数

$$\lambda g_1(t) + g_2(t) \tag{A-4}$$

其中λ为一实数变量，记

$$f(\lambda) = \int_a^b [\lambda g_1(t) + g_2(t)]^2 dt \tag{A-5}$$

$$= \lambda^2 \int_a^b |g_1(t)|^2 dt + 2\lambda \int_a^b g_1(t) g(t) dt + \int_a^b |g_2(t)|^2 dt \tag{A-6}$$

记
$$A = \int_a^b |g_1(t)|^2 dt \geqslant 0 \tag{A-7}$$

$$B = \int_a^b g_1(t) g_2(t) dt \tag{A-8}$$

$$C = \int_a^b |g_2(t)|^2 dt \geqslant 0 \tag{A-9}$$

则
$$f(\lambda) = \lambda^2 A + 2B\lambda + C \tag{A-10}$$

因为$f(\lambda)$是非负的，除了可能得到实数重根外不会有别的实根存在，即 $f(\lambda) = 0$ 的根一定取共轭对或实数重根，其条件应当满足

$$B^2 \leqslant AC \tag{A-11}$$

把(A-7)、(A-8)、(A-9)三式代入(A-11)式，得

$$\left|\int_a^b g_1(t) g_2(t) dt\right|^2 \leqslant \left(\int_a^b |g_1(t)|^2 dt\right)\left(\int_a^b |g_2(t)|^2 dt\right) \tag{A-12}$$

(A-12) 式通常称为施瓦兹不等式。

容易证明，当$g_1(t)$与$g_2(t)$成比例关系时，即有$g_1(t) = k g_2(t)$，k为任意常数，则式(A-12)取等号。

若$g_1(t)$、$g_2(t)$为复数函数，且$g_1(t)$、$g_2(t)$满足(A-1)、(A-2)式，则得到复函数的施若兹不等式如下

$$\left|\int_a^b [g_1^*(t) g_2(t) + g_1(t) g_2^*(t)] dt\right|^2 \leqslant 4\left(\int_a^b |g_1(t)|^2 dt\right)\left(\int_a^b |g_2(t)|^2 dt\right) \tag{A-13}$$

证明(A-13)的方法与前面相似，记

$$f(\lambda)=\int_a^b[\lambda g_1^*(t)+g_2^*(t)][\lambda g_1(t)+g_2(t)]dt$$

$$=\lambda^2 A+\lambda(B+B^*)+C \qquad (A\text{-}14)$$

其中

$$A=\int_a^b|g_1(t)|^2dt\geqslant 0 \qquad (A\text{-}15)$$

$$B=\int_a^b g_1^*(t)g_2(t)dt \qquad (A\text{-}16)$$

$$C=\int_a^b|g_2(t)|^2\geqslant dt\geqslant 0 \qquad (A\text{-}17)$$

$f(\lambda)$不能取负值的条件为

$$(B+B^*)^2\leqslant 4AC \qquad (A\text{-}18)$$

把(A-14)、(A-15)、(A-16)三式代入，得

$$\left|\int_a^b[g_1^*(t)g_2(t)+g_1(t)g_2^*(t)]dt\right|^2$$

$$\leqslant 4\left(\int_a^b|g_1(t)|^2dt\right)\left(\int_a^b|g_2(t)|^2dt\right)$$

特殊情况

若

$$g_1(t)=g_2(t)=\sqrt{g(t)} \qquad (A\text{-}19)$$

把(A-19)代入(A-3)式，可得

$$\left|\int_a^b g(t)dt\right|^2\leqslant\left[\int_a^b(g(t))dt\right]^2 \qquad (A\text{-}20)$$

或写成另一种等价形式

$$\left|\int_a^b g(t)dt\right|\leqslant\int_a^b|g(t)|dt \qquad (A\text{-}21)$$

所以，在函数$g(t)$曲线下的总面积的绝对值小于或等于$g(t)$绝对值的积分。

矢量施瓦兹不等式

设有矢量X、Y，则有矢量施瓦兹不等式

$$|X^TY|^2\leqslant(X^TX)(Y^TY) \qquad (A\text{-}22)$$

证：令λ为任意实数，有

$$(X+\lambda Y)^T(X+\lambda Y)$$

$$=X^TX+2\lambda X^TY+\lambda^2Y^TY\geqslant 0 \qquad (A\text{-}23)$$

若取

$$\lambda=-\frac{X^TY}{Y^TY} \qquad (A\text{-}24)$$

代入(A-23)式中，有

$$X^TX-\frac{2|X^TY|^2}{Y^TY}+\frac{|X^TY|^2}{Y^TY}\geqslant 0$$

或

$$|X^TY|^2\leqslant(X^TX)(Y^TY)$$

即如所证。

附录B　欠定方程组的最小范数解

设有线性方程组，其中$m<n$

$$
\begin{cases}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\
\cdots\cdots\cdots\cdots\cdots\cdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m
\end{cases}
\tag{B-1}
$$

(B-1) 式称为欠定方程组。其中方程数目少于未知量数目。写成矢量形式，得

$$AX = b \tag{B-2}$$

求未知量 x_1，x_2，…，x_n，的解。

因为未知量的数目大于方程组的数目，所以有无穷多组解满足方程组 (B-1)，但其中只有一组解最有意义，我们把满足

$$J = x_1^2 + x_2^2 + \cdots + x_n^2 \tag{B-3}$$

取极小值的解称为最小范数解，J称为矢量 X的范数，下面来求最小范数解。

上述问题实质上就是求函数

$$f(X) = X^T X \tag{B-4}$$

在$AX = b$约束条件下的极小值，采用拉格朗日乘子法，构造函数

$$F(X) = X^T X + \Lambda^T(AX - b) \tag{B-5}$$

其中Λ^T代表拉格朗日乘子，为m维行向量，上式对X求导数，并使之等于零，得

$$\frac{dF}{dX} = 2X + A^T\Lambda = 0 \tag{B-6}$$

解出

$$X = -\frac{1}{2}A^T\Lambda \tag{B-7}$$

将它代入约束方程，可得

$$-\frac{1}{2}AA^T\Lambda = b \tag{B-8}$$

其中AA^T是$m \times m$维常数矩阵，根据给定条件可知它的秩为m，其逆存在，因而有

$$\Lambda = -2(AA^T)^{-1}b \tag{B-9}$$

将它代入(B-7)式，便得到

$$X = A^T(AA^T)^{-1}b \tag{B-10}$$

又由

$$\frac{d}{dX^T}\frac{dF}{dX} = \frac{d^2F}{dX^2} = 2I > 0 \tag{B-11}$$

可知，(B-10)式代表最小范数解。

倘若在(B-1)中，$m > n$ 方程式数目多于未知量的数目；这时可采用第二章中所述的最小二乘法求解。

附录C 矩 阵 导 数

由于本书中经常用到矩阵导数，但在一般数学书籍或手册中均较少讲到，为方便起见，列出本附录以便查阅。

一、自变量为标量

1. 设有矢量 $a(t) = [a_1(t),\ a_2(t),\ \cdots,\ a_n(t)]^T$

则对 t 的导数为

$$\frac{d}{dt}a(t)=\left[\frac{da_1(t)}{dt},\ \frac{da_2(t)}{dt},\cdots,\frac{da_n(t)}{dt}\right]^T \tag{C-1}$$

2． 设有 $n\times m$ 矩阵

$$A(t)=\begin{pmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1m}(t) \\ a_{2}(t) & a_{21}(t) & \cdots & a_{2m}(t) \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}(t) & a_{n2}(t) & \cdots & a_{nm}(t) \end{pmatrix}$$

对 t 的导数为

$$\frac{dA(t)}{dt}=\begin{pmatrix} \frac{da_{11}}{dt} & \frac{da_{12}}{dt} & \cdots & \frac{da_{1m}}{dt} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{da_{n1}}{dt} & \frac{da_{n2}}{dt} & \cdots & \frac{da_{nm}}{dt} \end{pmatrix} \tag{C-2}$$

3． 设有矩阵 $A(t)$ 和 $B(t)$，则有

$$\frac{d}{dt}(A\pm B)=\frac{dA}{dt}\pm\frac{dB}{dt} \tag{C-3}$$

$$\frac{d}{dt}(AB)=\frac{dA}{dt}B+A\frac{dB}{dt} \tag{C-4}$$

4． 设 A 为 $n\times n$ 维对称常数矩阵，X 为 t 的 n 维矢量函数，则有

$$\frac{d}{dt}X^TAX=2X^TA\frac{dX}{dt} \tag{C-5}$$

二、自变量为矢量

1． 设有函数 $f(X)$，$X=[x_1,x_2,\cdots,x_n]^T$，则有

$$\frac{\partial f}{\partial X}=\begin{pmatrix} \frac{\partial f}{\partial X_1} \\ \frac{\partial f}{\partial X_2} \\ \vdots \\ \frac{\partial f}{\partial X_n} \end{pmatrix} \tag{C-6}$$

若 $f(X)=X^TX$，则有 $\frac{\partial f(X)}{\partial X}=2X$ (C-7)

2． 设有标量函数 $f(X)$ 和 $g(X)$，则有

$$\frac{d}{dX}(f\pm g)=\frac{df}{dX}\pm\frac{dg}{dX} \tag{C-8}$$

$$\frac{d}{dX}(f\cdot g)=\frac{df}{dX}g+f\frac{dg}{dX} \tag{C-9}$$

3. 设有函数$f(X)$，对X^T求导，则有

$$\frac{df}{dX^T}=\left[\frac{\partial f}{\partial X_1},\ \frac{\partial f}{\partial X_2},\cdots,\frac{\partial f}{\partial X_n}\right] \tag{C-10}$$

4. 设有矢量函数 $a(X)=\begin{pmatrix}a_1(X)\\a_2(X)\\\vdots\\a_m(X)\end{pmatrix}$，则有

$$\frac{d\,a^T(X)}{dX}=\begin{pmatrix}\frac{\partial a_1}{\partial x_1} & \frac{\partial a_2}{\partial x_1} & \cdots & \frac{\partial a_m}{\partial x_1}\\ \frac{\partial a_1}{\partial x_2} & \frac{\partial a_2}{\partial x_2} & \cdots & \frac{\partial a_m}{\partial x_2}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial a_1}{\partial x_n} & \frac{\partial a_2}{\partial x_n} & \cdots & \frac{\partial a_m}{\partial x_n}\end{pmatrix} \tag{C-11}$$

$$\frac{d\,a(X)}{dX^T}=\begin{pmatrix}\frac{\partial a_1}{\partial x_1} & \frac{\partial a_1}{\partial x_2} & \cdots & \frac{\partial a_1}{\partial x_n}\\ \frac{\partial a_2}{\partial x_1} & \frac{\partial a_2}{\partial x_2} & \cdots & \frac{\partial a_2}{\partial x_n}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial a_m}{\partial x_1} & \frac{\partial a_m}{\partial x_2} & \cdots & \frac{\partial a_m}{\partial x_n}\end{pmatrix} \tag{C-12}$$

和

$$\frac{d\,a^T(X)}{dX}=\left[\frac{d\,a(X)}{dX^T}\right]^T \tag{C-13}$$

$$\frac{dX^T}{dX}=1 \tag{C-14}$$

$$\frac{dX}{dX^T}=1 \tag{C-15}$$

5. 设$a(X)$和$b(X)$是m维的列矢量函数，则有

$$\frac{d}{dX}(a^T\pm b^T)=\frac{d\,a^T}{dX}\pm\frac{d\,b^T}{dX} \tag{C-16}$$

$$\frac{d}{dX}(a^T b)=\frac{d\,a^T}{dX}b+\frac{d\,b^T}{dX}a \tag{C-17}$$

$$\frac{d}{dX}(\lambda a^T)=\frac{d\lambda}{dX}a^T+\lambda\frac{d\,a^T}{dX}\quad(\lambda=\lambda(X)\text{为标量函数}) \tag{C-18}$$

6. 设A、B为常数矩阵，则有

$$\frac{d}{dX}(X^T A)=A \tag{C-19}$$

$$\frac{d}{d X^T}(B X)=B \tag{C-20}$$

$$\frac{d}{d X}(X^T A X)=(A+A^T) X \tag{C-21}$$

$$\frac{d}{d X}(P^T A X)=A^T P \qquad (P^T\text{为}1\times n\text{行矢量}) \tag{C-22}$$

7．设有矩阵函数

$$A(X)=\begin{pmatrix} a_{11}(X) & \cdots & a_{1l}(X) \\ \cdots & \cdots & \cdots \\ a_{m1}(X) & \cdots & a_{ml}(X) \end{pmatrix}$$

则有

$$\frac{d A(X)}{d X}=\begin{pmatrix} \frac{\partial A}{\partial x_1} \\ \frac{\partial A}{\partial x_2} \\ \vdots \\ \frac{\partial A}{\partial x_n} \end{pmatrix} \tag{C-23}$$

$$\frac{d A(X)}{d X^T}=\left[\frac{\partial A(X)}{\partial x_1},\ \frac{\partial A(X)}{\partial x_2},\cdots,\frac{\partial A(X)}{\partial x_n}\right] \tag{C-24}$$

$$\frac{d A^T}{d X}=\left[\frac{d A}{d X^T}\right]^T \tag{C-25}$$

8．设$A(X)$、$C(X)$都是$p\times m$矩阵，$B(X)$是$m\times l$矩阵，λ是X的标量函数，则有

$$\frac{d}{d X}(A\pm C)=\frac{d A}{d X}\pm\frac{d C}{d X} \tag{C-26}$$

$$\frac{d}{d X}(\lambda A)=\frac{d\lambda}{d X}\otimes A+\lambda\frac{d A}{d X} \tag{C-27}$$

其中$\otimes$表示克罗内克积。

$$\frac{d\lambda}{d X}\otimes A=\begin{pmatrix} \frac{\partial\lambda}{\partial x_1}A \\ \frac{\partial\lambda}{\partial x_2}A \\ \vdots \\ \frac{\partial\lambda}{\partial x_n}A \end{pmatrix}$$

$$\frac{d}{d X}(AB)=\frac{d A}{d X}B+A\otimes\frac{d B}{d X} \tag{C-28}$$

其中

$$A\otimes\frac{dB}{dX}=\begin{pmatrix}A\frac{\partial B}{\partial x_1}\\ A\frac{\partial B}{\partial x_2}\\ \vdots\\ A\frac{\partial B}{\partial x_n}\end{pmatrix}$$

三、自变量为矩阵

1. 设自变量为矩阵A，$f(A)$为A的标量函数，并设A为$p\times m$维矩阵，则有

$$\frac{df}{dA}=\begin{pmatrix}\frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} & \cdots & \frac{\partial f}{\partial a_{1m}}\\ \frac{\partial f}{\partial a_{21}} & \frac{\partial f}{\partial a_{22}} & \cdots & \frac{\partial f}{\partial a_{2m}}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial f}{\partial a_{p1}} & \frac{\partial f}{\partial a_{p2}} & \cdots & \frac{\partial f}{\partial a_{pm}}\end{pmatrix} \tag{C-29}$$

2. 设$f=X^TAX$，A为对称矩阵，则有

$$\frac{d}{dA}(X^TAX)=XX^T \tag{C-30}$$

3. 设A有的矢量函数$Z(A)$

$$Z(A)=\begin{pmatrix}z_1(A)\\ z_2(A)\\ \vdots\\ z_n(A)\end{pmatrix}$$

其中A为$p\times m$维矩阵，则$Z(A)$为$np\times m$维矩阵，于是有

$$\frac{dZ}{dA}=\begin{pmatrix}\frac{\partial Z}{\partial a_{11}} & \cdots & \frac{\partial Z}{\partial a_{1m}}\\ \cdots & \cdots & \cdots\\ \frac{\partial Z}{\partial a_{p1}} & \cdots & \frac{\partial Z}{\partial a_{pm}}\end{pmatrix} \tag{C-31}$$

其中

$$\frac{\partial Z}{\partial a_{ij}}=\begin{pmatrix}\frac{\partial Z_1}{\partial a_{ij}}\\ \frac{\partial Z_2}{\partial a_{ij}}\\ \vdots\\ \frac{\partial Z_n}{\partial a_{ij}}\end{pmatrix}$$

同理

$$\frac{dZ^T(A)}{dA}=\begin{pmatrix}\frac{\partial Z^T}{\partial a_{11}} & \cdots & \frac{\partial Z^T}{\partial a_{1m}}\\ \cdots & \cdots & \cdots\\ \frac{\partial Z^T}{\partial a_{p1}} & \cdots & \frac{\partial Z^T}{\partial a_{pm}}\end{pmatrix} \tag{C-32}$$

其中

$$\frac{\partial Z^T}{\partial a_{ij}}=\left[\frac{\partial Z_1^T}{\partial a_{ij}},\ \frac{\partial Z_2^T}{\partial a_{ij}},\cdots,\frac{\partial Z_n^T}{\partial a_{ij}}\right]$$

4．设函数

$$F(A)=\begin{bmatrix} f_{11}(A) & \cdots & f_{1l}(A) \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ f_{n1}(A) & \cdots & f_{nl}(A)\end{bmatrix}$$

是$n\times l$维矩阵函数，自变量A为$p\times m$维矩阵，则有

$$\frac{dF}{dA}=\begin{pmatrix} \frac{\partial F}{\partial a_{11}} & \cdots & \frac{\partial F}{\partial a_{p1}} \\ \cdots\cdots & \cdots & \cdots\cdots \\ \frac{\partial F}{\partial a_{1m}} & \cdots & \frac{\partial F}{\partial a_{pm}}\end{pmatrix} \tag{C-33}$$

5．设A是$n\times m$矩阵，X是n维列矢量，Y是m维列矢量，则有

$$\frac{\partial}{\partial A}(X^TAY)=XY^T \tag{C-34}$$

$$\frac{\partial}{\partial A}(Y^TA^TX)=XY^T \tag{C-35}$$

$$\frac{\partial}{\partial A}(Y^TA^TAY)=2AYY^T \tag{C-36}$$

$$\frac{\partial}{\partial A}(AY-X)^T(AY-X)=2(AY-X)Y^T \tag{C-37}$$

由于 $(AY-X)^T(AY-X)=\mathrm{tr}(AY-X)(AY-X)^T$，所以又得出

$$\frac{\partial}{\partial A}\mathrm{tr}(AY-X)(AY-X)^T=2(AY-X)Y^T \tag{C-38}$$

$$\frac{\partial}{\partial A}(ACA^T)=A(C+C^T) \tag{C-39}$$

$$\frac{\partial}{\partial A}(X^TA^TAX)=2AXX^T \tag{C-40}$$

四、复合函数的导数

我们用f代表标量函数，X代表n维列矢量或列矢量函数，Y代表m维列矢量或列矢量函数，Z代表l维列矢量函数，t代表标量自变量，我们有

1．设$f=f(Y)$，$Y=Y(t)$，则有

$$\frac{df}{dt}=\frac{df}{dY^T}\ \frac{dY}{dt}=\frac{dY^T}{dt}\ \frac{df}{dY} \tag{C-41}$$

2．设$f=f(Y)$，$Y=Y(X)$，则有

$$\frac{df}{dX}=\frac{dY^T}{dX}\cdot\frac{df}{dY} \tag{C-42}$$

$$\frac{df}{dX^T}=\frac{df}{dY^T}\cdot\frac{dY}{dX^T} \tag{C-43}$$

3. 设$f=f(X,Y),Y=Y(X)$，则有

$$\frac{df}{dX}=\frac{\partial f}{\partial X}+\frac{dY^T}{dX}\frac{\partial f}{\partial Y} \tag{C-44}$$

$$\frac{df}{dX^T}=\frac{\partial f}{\partial X^T}+\frac{\partial f}{\partial Y^T}\frac{dY}{dX^T} \tag{C-45}$$

$$\frac{d}{dX}(X^TAX)=(A+A^T)X \qquad \text{(其中}A\text{为常数矩阵)} \tag{C-46}$$

$$\frac{d}{dX}(AX-b)^T(AX-b)=2A^T(AX-b) \tag{C-47}$$

$$\frac{\partial}{\partial a}(X-a-BZ)^T(X-a-BZ)$$

$$=-2(X-a-BZ) \tag{C-48}$$

4. 设$Z_1=Z_1(Y)$， $Y=Y(t)$，则有

$$\frac{dZ}{dt}=\frac{dZ}{dY}\frac{dY}{dt} \tag{C-49}$$

5. 设$Z=Z(Y),Y=Y(X)$，则有

$$\frac{dZ^T}{dX}=\frac{dY^T}{dX}\frac{dZ^T}{dY} \tag{C-50}$$

$$\frac{dZ}{dX^T}=\frac{dZ}{dX^T}\frac{dY}{dX^T} \tag{C-51}$$

6. 设 $Z=Z(X,Y),Y=Y(X)$，则有

$$\frac{dZ^T}{dX}=\frac{\partial Z^T}{\partial X}+\frac{dY^T}{dX}\frac{\partial Z^T}{\partial Y} \tag{C-52}$$

$$\frac{dZ}{dX^T}=\frac{\partial Z}{\partial X^T}+\frac{\partial Z}{\partial Y^T}\frac{dZ}{dX^T} \tag{C-53}$$

$$\frac{d}{dX}(AX-b)^TR(AX-b)$$

$$=A^T(R+R^T)(AX-b) \tag{C-54}$$

五、矩阵的迹与行列式的导数

设X为$n\times n$维矩阵。

1. $\frac{\partial}{\partial X}\mathrm{tr}[X]=I$ (C-55)

2. $\frac{\partial}{\partial X}\mathrm{tr}[AX]=A^T$ (C-56)

3. $\frac{\partial}{\partial X}\mathrm{tr}[AX^T]=A$ (C-57)

4. $\frac{\partial}{\partial X}\mathrm{tr}[AXB]=A^TB^T$ (C-58)

5. $\frac{\partial}{\partial \boldsymbol{X}}\mathrm{tr}[\boldsymbol{A}\boldsymbol{X}^T\boldsymbol{B}]=\boldsymbol{B}\boldsymbol{A}$ (C-59)

6. $\frac{\partial}{\partial \boldsymbol{X}}\mathrm{tr}[\boldsymbol{X}\boldsymbol{X}]=2\boldsymbol{X}^T$ (C-60)

7. $\frac{\partial}{\partial \boldsymbol{X}}\mathrm{tr}[\boldsymbol{X}\boldsymbol{X}^T]=2\boldsymbol{X}$ (C-61)

8. $\frac{\partial}{\partial \boldsymbol{X}}\mathrm{tr}[\boldsymbol{X}^n]=n[\boldsymbol{X}^{n-1}]^T$ (C-62)

9. $\frac{\partial}{\partial \boldsymbol{X}}\mathrm{tr}[\boldsymbol{A}\boldsymbol{X}^n]=\left[\sum_{i=0}^{n-1}\boldsymbol{X}^i\boldsymbol{A}\boldsymbol{X}^{n-1-i}\right]^T$ (C-63)

10. $\frac{\partial}{\partial \boldsymbol{X}}\mathrm{tr}[\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}]=\boldsymbol{A}^T\boldsymbol{X}^T\boldsymbol{B}^T+\boldsymbol{B}^T\boldsymbol{X}^T\boldsymbol{A}^T$ (C-64)

11. $\frac{\partial}{\partial \boldsymbol{X}}\mathrm{tr}[\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}\boldsymbol{X}^T]=\boldsymbol{A}^T\boldsymbol{X}\boldsymbol{B}^T+\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}$ (C-65)

12. $\frac{\partial}{\partial \boldsymbol{X}}\mathrm{tr}[\boldsymbol{X}^{-1}]=(\boldsymbol{X}^{-1}\boldsymbol{X}^{-1})^T=-(\boldsymbol{X}^{-2})^T$ (C-66)

13. $\frac{\partial}{\partial \boldsymbol{X}}\mathrm{tr}[\boldsymbol{A}\boldsymbol{X}^{-1}\boldsymbol{B}]=-(\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{A}\boldsymbol{X}^{-1})^T$ (C-67)

14. $\frac{\partial}{\partial \boldsymbol{X}}\det[\boldsymbol{X}]=[\det(\boldsymbol{X})][\boldsymbol{X}^{-1}]^T$ (C-68)

15. $\frac{\partial}{\partial \boldsymbol{X}}\det[\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}]=(\det[\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}])(\boldsymbol{X}^{-1})^T$ (C-69)

16. $\frac{\partial}{\partial \boldsymbol{X}}\det[\boldsymbol{X}^T]=\frac{\partial}{\partial \boldsymbol{X}}\det(\boldsymbol{X})$ (C-70)

17. $\frac{\partial}{\partial \boldsymbol{X}}\det[\boldsymbol{X}^n]=n[\det(\boldsymbol{X})]^n(\boldsymbol{X}^{-1})^T$ (C-71)

18. $\frac{\partial}{\partial \boldsymbol{X}}\ln(\det[\boldsymbol{X}])=(\boldsymbol{X}^{-1})^T$ (C-72)

附录D 五个常用的矩阵等式

1. $(\boldsymbol{P}^{-1}+\boldsymbol{H}^T\boldsymbol{Q}^{-1}\boldsymbol{H})\boldsymbol{P}\boldsymbol{H}^T=\boldsymbol{H}^T\boldsymbol{Q}^{-1}(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\boldsymbol{Q})$ (D-1)

证：把两端括号打开，得

$$\boldsymbol{H}^T+\boldsymbol{H}^T\boldsymbol{Q}^{-1}\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T=\boldsymbol{H}^T\boldsymbol{Q}^{-1}\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\boldsymbol{H}^T$$

即如所证。

2. $\boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\boldsymbol{Q})^{-1}=(\boldsymbol{P}^{-1}+\boldsymbol{H}^T\boldsymbol{Q}^{-1}\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{Q}^{-1}$ (D-2)

证：用$(\boldsymbol{P}^{-1}+\boldsymbol{H}^T\boldsymbol{Q}^{-1}\boldsymbol{H})$左乘上式，再用$(\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T+\boldsymbol{Q})$右乘上式，得

$$H^T + H^T Q^{-1} HPH^T = H^T Q^{-1} HPH^T + H^T$$

即如所证。

3. $(P^{-1} + H^T R^{-1} H)^{-1} = P - PH^T(HPH^T + R)^{-1} HP$ (D-3)

证：用$P^{-1} + H^T R^{-1} H$左乘上式，得

$$I = (P^{-1} + H^T R^{-1} H)[P - PH^T(HPH^T + R^{-1})^{-1} HP]$$

再利用(D-2)式，得上式右边等于I，

即如所证。

4. $(I + PH^T R^{-1} H)^{-1} P = (P^{-1} + H^T R^{-1} H)^{-1}$ (D-4)

证：用$P^{-1} + H^T R^{-1} H$右乘上式，得

$$(I + PH^T R^{-1} H)^{-1} P(P^{-1} + H^T R^{-1} H) = I$$

即如所证。

5. $H(P^{-1} + H^T R^{-1} H)^{-1} H^T = R - R(HPH^T + R)^{-1} R$ (D-5)

证：由(D-4)式，得

$$\begin{aligned} H(P^{-1} + H^T R^{-1} H)^{-1} H^T &= HPH^T(HPH^T + R)^{-1} R \\ &= (HPH^T + R - R)(HPH^T + R)^{-1} R \\ &= R - R(HPH^T + R)^{-1} R \end{aligned}$$

即如所证。

参 考 文 献

第一篇 变 量 估 计

[1] T.Kailath, A View of Three Decades of Linear Filtering Theory, IEEE. Trans. Inform. Theory, Vol IT-20, No2, March 1974.

[2] H.W.Sorenson, Least-square Estimation, From Ganss to Kalman, IEEE, Tran, Spectrum, Vo17, No2, July 1979.

[3] H.L.范特里斯，毛士艺等译,检测、估计和调制理论,国防工业出版社，1983.

[4] A.D.惠伦，刘其培等译，噪声中信号的检测，科学出版社，1977.

[5] M.D.斯里纳思、P.K.雷杰斯卡兰，朱正中等译，统计信号处理及其应用导论，国防工业出版社，1982。

第二篇 维 纳 滤 波

[1] 何振亚，数学信号处理的理论与应用，人民邮电出版社，1983.

[2] 鞠德航等，信号检测理论导论，科学出版社，1977.

[3] N.阿罕默得、K.R.罗，胡正名等译，数字信号处理中的正交变换，人民邮电出版社，1979.

[4] 郑兆宁、向大威，水声信号被动检测与参数估计理论，科学出版社，1983.

[5] B.Widrow, et.al, Adaptive Noise Cancelling Principle and Application, Proc IEEE, Vol63, Dec 1975,

[6] B. Widrow, et.al,Stationary and Nonstationary Learning Characteristics of The LMS Adaptive Filter, Proc, IEEE, Vol4, No8, Aug 1976.

[7] D.C. Parden, et.al, On The Desired Behavior of Adaptive Signal Processing Algorithm, Inf, Conf, on ASSP, 1979

[8] D. C. Parden, et.al. Tracking Properties of Adaptive Processing Algorithm, Int. Conf. on ASSP, 1979.

[9] O.L.Frost, An Algorithm for Lineary-Constrained Adaptive Array Processing, PIEEE, 60, Aug.1972.

[10]许树声，信号检测与估计，国防工业出版社，1985.

第三篇 卡尔曼滤波

[1] 中国科学院数学研究所概率组，离散时间系统滤波的数学方法，国防工业出版社，1975.

[2] 贾沛璋、朱征桃，最优估计及其应用，科学出版社，1984.

[3] 谢绪恺，现代控制理论基础，辽宁人民出版社，1980.

[4] 孙仲康，雷达数据数字处理，国防工业出版社，1983.

[5] 邓自立、郭一新，动态系统分析及其应用，辽宁科学技术出版社，1985.

[6] B.D.O.安德森、J.B.摩尔，卢伯英译，最佳滤波，国防工业出版社，1983.

[7] 周宏仁，机动目标当前统计模型与自适应跟踪算法，航空学报，1983.

[8] A. H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press. Inc.1970

[9] P.S. Mayoeck, Stochastic Models, Estimation and Control, Academic Press. Inc,1982.

[10] K. K. Mehra, Approaches to Adaptive Filtering, IEEE. Trans, Voll, AC-17,No5, 1972.

[11] M. M. Sandhi, D. Mehsn, New Results on The Performance of A Well-Known Class of Adaptive Filters, Proc, IEEE, Vol 63, Noll, Nov.1976.

[12] K.A.Myers, B. D. Tapley, Adaptive Sequential Estimation With Unknown Noise Statistics, IEEE, Trans, Vol AC-21,No4,1978.

[13] K.Tajima, Estimation of Steady-state Kalman Filter Gain,IEEE. Trans, Vol AC-23, No5, 1978.

[14] Y. T. Chen, A.G.C.Hu, J. B. Plant, A Kalman Filter Tracking Scheme With Input Estimation, IEEE, Trans, Vol AES-15, No2, March 1979.

[15] Y. T. Chen, J. B. Plant, A Kalman Tracker With Simple Input Etimator, IEEE, Trans, Vol AES-18, No2, March 1982

[16] N.K. Sinha, A.F.W. Tom, Adaptive State Estimation For Systems With Coloured Observation Noise, J.Syst. Sci, Vol9, Noll, 1978.

[17] A. P. 塞奇、J.L.梅尔萨，田承骏等译，估计理论及其在通讯与控制中的应用，科学出版社，1977.

[18] 袁天鑫，最佳估计原理，国防工业出版社，1980.

[19] Xiang Chu Qi, Wei Wei, A Kalman Filter Used for Passive Ranging, IEEE.1988 International Workshop on Infoamation Theory, Beijing, China

[20] 项楚骐、魏伟，自适应卡尔曼滤波的新算法，船工学报，1988，第三期。

[21] 项楚骐，卡尔曼滤波的存在性自动化技术与应用，1988，第四期。

[22] 项楚骐，水下噪声源定位，海洋技术，1988，第二期。

[23] 项楚骐、杨椿，水下噪声源跟踪与自适应卡尔曼滤波，海洋技术，1988，第三期。

[24] Xiang Chu Qi, The Underwater Noise Source Located by Passive Sonar System, WESTRAC III, 1988, Shanghai, China.

[1]. J. P. Burg. 'Maximum Entropy Spectral Analysis', 37th Ann. Intern. Meeting, Soc. of Explor. Geophys, Oklahoma City, Oklahoma(1967).

[2]. Van den Bos, 'Alternative Interpretation of Maximum Entropy Spectral Analysis', IEEE Trans. Inform. Theory, IT-17, pp. 493—494,1971.

[3]. J. Capon, 'High Resolution Frequency-wave-number Spectrum Analysis', Proc. IEEE, Vol. 57, pp. 1408—1418,1969.

[4]. R. T. Lacoss, 'Data Adaptive Analysis Methods', Geophysics, Vol.36,No.4,pp.661—675,1971.

[5]. S. Treitel and E. A. Robinson, 'The Design of High-resolution Digital Filters', IEEE Trans. Geosci, Electronic, GE-4 No.1,pp. 25—38,1966.

[6]. J. Makhuol, 'Linear Prediction: A Tutoral Review', Proc. IEEE, Vol. 63, No.4,pp.561—580, 1975.

[7]. A.J.Berkhout, 'On the Minimum Phase Criterion of Sampled Signals', IEEE Trans. Geosci, Electromics,GE-11,pp.186—198,1973.

[8]. T.J.Ulrych, R.W.Clayton, 'Time Series Modelling and Maximum Entropy',physics of earth and planetary Interiors, 12,pp.188—200, 1976.

[9]. L.J.Griffiths,'Rapid Measurement of Digital Instantaneous Frequency', IEEE Trans. ASSP, Vol. ASSP-23, No.2,pp,207—222,1975.

[10]. H. Akaike, 'Fitting Autoregressive Models for Prediction', Ann. Inst. Math., Vol.21, pp.243—247, 1969.

[11]. H. Akaike, 'A New Look at the Statistical Model Identification', IEEE Trans, Auton. Control, Vol. AC-19, pp. 716—723, 1974.

[12]. Nonlinear Methods of Spetral Analysis,Edited by S.Haykin, 1979.

[13]. E. Parzen, 'An Approach to Time Series Modeling and Forecasting Illustrated by Hourly Electricity Demands'. Statistical Sci. Div.,State University of New York in Buffalo,Tech.Rpt.37(1976).

[14]. J. P. Burg, 'A New Analysis Technique for Time Series Data', Presented at Advanced Study Institute on Signal Processing,NATO, Enschede, Netherlands, 1968.

[15]. D.E. Smylie, G. K. G. Clarke, and T. J. Ulrych, 'Analysis of Irregularities in the Earth's Rotation', in Methods In Computational Physics, Vol.13,pp.391—430, 1973.

[16]. J. E. Shore and R. W. Johnson, 'Axiomatic Derivation of the

Princinciple of Maximum Cross-Entropy', IEEE Trans. on Inform, Theory, Vol.IT-26, No.1, Janary 1980.

[17]. J. E. Shore, 'Minimum Cross-Entropy Spectral Analysis', IEEE Trans. on Acoustics, speech and signal processing, Vol. ASSP-29, No.2, April 1981.

[18]. J.E. Shore and R.W. Johnson,'Proprties of Cross-Entropy Minimization', IEEE Trans. Inform. Theory Vol. JT-27.1981.

[19]. H.S.Konvalenka etal., 'Simaltaneus Estimation of Poles and Zeros in Speech Analysis and ITIF-Iterative Inverse Fitting Aigorithm', IEEE Trans. Acous., Speech, Signal Process., pp200—202, Apr.1979.

[20]. A. Papoulis, Signal Analysis. New York: Mc Graw-Hill, 1977, p231.

[21]. 何振亚编著，数字信处理的理论与应用。人民邮电出版社，1983年12月.

[22]. R.B. Blackman and J.W.Tukey, The Measurement of Power Spectra, Dover Publications Inc. New York, 1958.

[23]. M. Schwartz, L. Shaw, Signal Processing, Discrete Spectral Analysis, Detection, and Estimation. Mc Graw-Hill, 1675.

[24]. A.V.Oppenheim, R.W.Schafer, Digital Signal Processing. prentics-Hall,1975.（中译本：数字信号处理。科学出版社,1981）

[25]. P.D.Welch, 'The Use of Fast Fourier Transform for the Estimation of Power Spectra', IEEE. Trans.Audio Electroacoust., Vol AU-15, June 1979